(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 November 2001 (29.11.2001)

PCT

(10) International Publication Number
WO 01/90343 A2

(51) International Patent Classification[7]: C12N 15/00

(21) International Application Number: PCT/US01/16659

(22) International Filing Date: 22 May 2001 (22.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/206,136 22 May 2000 (22.05.2000) US

(71) Applicant *(for all designated States except US)*: **COLD SPRING HARBOR LABORATORY** [US/US]; One Bungtown Road, Cold Spring Harbor, NY 11724 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **MARTIENSSEN, Robert, A.** [US/US]; One Bungtown Road, Cold Spring Harbor, NY 11724 (US). **VOLLBRECHT, Erik** [US/US]; One Bungtown Road, Cold Spring Harbor, NY 11724 (US).

(74) Agent: **NEBEL, Heidi, S.**; Zarley, McKee, Thomte, Voorhees & Sease, Suite 3200, 801 Grand Avenue, Des Moines, IA 50309-2721 (US).

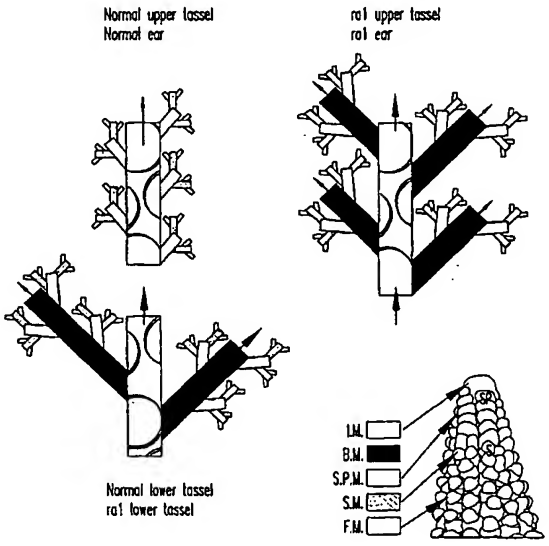
(81) Designated States *(national)*: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NUCLEOTIDE SEQUENCES ENCODING RAMOSA 1 GENE AND METHODS OF USE FOR SAME



(57) Abstract: The invention relates to the isolation and characterization of a novel maize gene (*Ra1*) responsible for meristem development and inflorescence development including branching. The novel gene, gene product, and regulatory regions may be used to manipulate branching, meristem growth, inflorescence development and arrangement, and ultimately to improve yield of plants. The invention includes the novel gene and protein product as well as the use of the same for temporal and spatial expression in transgenic plants to alter plant morphology and affect yield in plants.

WO 01/90343 A2

TITLE: NUCLEOTIDE SEQUENCES ENCODING RAMOSA 1 GENE AND METHODS OF USE FOR SAME

FIELD OF THE INVENTION

This invention relates generally to the field of plant molecular biology. More specifically, this invention relates to the characterization of a novel maize Ramosa 1 protein and a nucleotide sequence encoding the same as well as genetic techniques using the same for modification of plant architecture to increase yield and health of plants.

BACKGROUND OF THE INVENTION

Organogenesis in flowering plants occurs in meristems, which are relatively undifferentiated tissues located at the growing points of the plant (Steeves and Sussex 1989). Shoot meristems generate shoot components such as leaves, stems, flowers and branches, while root meristems generate the primary epidermal, cortical and vascular tissues of the root (Martienssen and Dolan 1998). In each case, the meristem appears to be compartmentalized into stem cells, which divide relatively slowly and retain stem cell identity, and their daughters, which can divide more rapidly and ultimately differentiate. In shoots, stem cells populate the central zone of the apex while their daughters populate the peripheral zone. Thus, shoot meristem function consists of self-renewal (which occurs in the central zone) and the coordinated production of organ and branch primordia (which occurs in the peripheral zone).

The determinacy of the meristem defines its relative capacity for self-renewal: Indeterminate meristems, such as vegetative shoot meristems, continue to produce leaves and branches in the absence of a floral stimulus, and so maintain a potentially unlimited capacity for organogenesis. In contrast, determinate meristems, such as floral meristems, initiate a fixed number of organ primordia before the stem cells differentiate into the final form (Sussex 1989).

Meristem determinacy ultimately dictates the architecture of the growing plant, specifying the arrangement and number of lateral organs and branches. It has an enormous impact on yield by altering the numbers of fruits and seeds produced by the inflorescences (due to extra branches) or by making plants more compact allowing them to be grown under stringent conditions (e.g. planted at high density or under adverse weather conditions). Thus, understanding how meristem determinacy is regulated is the key to understanding the diversity of plant form. Over the last few years, genes that regulate this process have been isolated from a variety of plants, including maize, *Arabidopsis*, tomato and *Antirrhinum*. These genes fall into two fundamental categories, those that determine meristem maintenance and those that determine meristem identity. This invention concerns the Ramosa 1 gene in maize, which regulates meristem identity in the inflorescence resulting in the stereotypical pattern of branching found in the tassel and in the ear. The pattern of meristem identity in the maize inflorescence is complex, providing an ideal system for uncovering and manipulating the genetic hierarchy that controls plant form. It also reflects the diversity found among related species in the grass family.

Meristem maintenance. In *Arabidopsis*, *SHOOT MERISTEMLESS* (*STM*), *PINHEAD/ZWILLE* and *WUSCHEL* are required to maintain the stem cell population in the shoot apex (Barton and Poethig 1993; Laux *et al.* 1996; McConnell and Barton 1995; Moussian *et al.* 1998). Similarly, the homeobox gene *knotted1* (*kn1*), a homolog of *STM*, is required for meristem maintenance in maize (Kerstetter *et al.* 1997; Vollbrecht *et al.* 1991)(E. Vollbrecht, L. Reiser and S. Hake, submitted). Conversely, in the *clavata* (*clv1* and *clv3*) mutants of *Arabidopsis*, stem cells fail to differentiate and meristems enlarge (Clark, Running and Meyerowitz 1993; Clark, Running and Meyerowitz 1995, Laufs et al., 1998; Leyser and Furner 1992). *clv1* and *stm* mutually suppress each other, and strong alleles of *stm* are epistatic to *clv1* mutations. Thus *CLV1* and *STM*, which encode a receptor kinase and a

homeodomain transcription factor respectively, likely function through opposing effects on a single process that specifies stem cell fate (Barton and Poethig 1993; Clark *et al.* 1996).

CLAVATA1 and *STM* also have opposite roles in the flower. *clv1* flowers display extensive fasciation, while in weak alleles of *stm* the floral meristem terminates early, with the loss of reproductive organs, especially the gynoecium. *STM* and *CLV* genes thus function in all shoot meristems, suggesting the existence of a signaling pathway regulating global shoot meristem structure.

Meristem identity. The primary shoot apical meristem adopts a series of identities as it progresses through distinct developmental phases (Allsopp 1967; Poethig 1990; Telfer and Poethig 1998). At the onset of the reproductive phase, a vegetative meristem converts into an indeterminate reproductive meristem, which in turn gives rise to determinate floral meristems arranged in a branching system called the inflorescence. In these transitions, meristem function is altered in response to environmental and developmental cues through meristem identity genes. These genes regulate meristem determinacy and the types of organs initiated. Thus both meristem maintenance and identity genes regulate the shoot meristem's capacity for self-renewal, suggesting members of these two classes of genes may interact.

In *Antirrhinum,* loss-of-function mutations in *FLORICAULA* (*FLO*), which encodes a presumed transcription factor, produce indeterminate inflorescence shoots in place of flowers (Coen *et al.* 1990). The same phenotype occurs when *Arabidopsis* plants are doubly mutant, for the *FLO* ortholog *LEAFY* and for *APETALA1*(*AP1*), suggesting these genes can act redundantly to specify meristem fate (Huala and Sussex 1992; Weigel *et al.* 1992). This view has recently been refined, as triple mutants of *AP1* and the *AP1* homologs *CAULIFLOWER* and *FRUITFULL* result in the complete conversion of flowers to shoots, suggesting that all three genes regulate LEAFY (Ferrandiz *et al.* 2000; Martienssen and Dolan 1998).

Loss-of-function mutations in *CENTRORADIALIS* (*CEN*) condition the reverse phenotype in *Antirrhinum*: a determinate inflorescence with only a few flowers (Bradley *et al.* 1996). In contrast, mutations of a *CEN* ortholog in tomato only diminish indeterminacy of sympodial renewal shoots without impacting identity of the inflorescence meristem (Pneuli *et al.* 1998). Thus, *CEN* specifies meristem indeterminacy to varying degrees in the inflorescence of these two species. In *Arabidopsis*, mutations of the *CEN* ortholog, *TERMINAL FLOWER 1 (TFL1)*, truncate the vegetative phase and abolish inflorescence meristem identity and indeterminacy (Shannon and Meeks-Wagner 1991). Gain and loss of function studies of *TFL1* suggest that, rather than simply negatively regulating *LFY* and *AP1*, *TFL1* acts on a central mechanism that regulates meristem identity throughout development. This is consistent with its proposed function as a signaling molecule in the inflorescence meristem (Bradley *et al.* 1997; Liljegren *et al.* 1999; Ratcliffe *et al.* 1998). These phenotypes suggest that a common mechanism underlies determinacy in these dicots, and that modifications of this *CEN*-based mechanism reflect diverse dicot inflorescence architectures.

In *Arabidopsis*, *APETALA2* is usually interpreted as encoding an organ identity gene responsible for specifying the fate of sepals and petals. However, it is expressed in the inflorescence meristem and may have a role in floral meristem establishment (Jofuku *et al.* 1994; Okamuro *et al.* 1997). Interestingly, ectopic *AP2* expression in petunia has drastic effects on inflorescence architecture, though the molecular mechanism is far from clear (Maes, Van Montagu and Gerats 1999).

The maize inflorescence: meristem identity and the determinacy series. The molecular genetics of meristem identity genes in maize is less well understood. Axillary meristems in the vegetative phase give rise to tillers, which reiterate development of the shoot and are tipped by branched tassels, though these are often feminized. In the reproductive phase, internode elongation in axillary shoots is drastically curtailed, and they are topped with

female inflorescences instead, which give rise to unbranched ears. Loss-of-function mutations in *teosinte branched 1* (*tb1*) affect axillary meristem determination in the reproductive phase, resulting in tiller-like shoots in place of ears. *tb1* encodes a TCP transcription factor which may act primarily by inhibiting axillary shoot growth (Cubas *et al.* 1999; Doebley, Stec and Hubbard 1997; Martienssen 1997).

In normal plants, when the floral stimulus is generated, development of first the tassel and then the ears commences from terminal and axillary meristems respectively (Lejeune and Bernier 1996). Initially, the pattern of meristem identity and inflorescence development in the tassel and ear are remarkably similar (Cheng, Greyson and Walden 1983) (Fig. 1). The main inflorescence meristem in each case is indeterminate, acropetally initiating a few hundred second-order meristems.

In the tassel, second order meristems assume either of two opposing fates: a few early-initiated ones become indeterminate branch meristems, and the rest become determinate spikelet pair meristems. In turn, branch meristems produce axes that initiate determinate spikelet pair meristems. In the ear, all second order meristems become determinate spikelet pair meristems.

Third order, spikelet meristems are similarly determinate in both inflorescences, and produce two fourth order, floral meristems. Floral meristems are the most determinate, producing only floral organs. This gradual but predictable progression through a series of switchpoints makes the maize inflorescence particularly useful for studying meristem determinacy (Postlethwait and Nelson 1964).

Mutations in maize inflorescence architecture. More than 30 mutations have been described that affect development of the maize inflorescence, but only a handful have been cloned molecularly (Veit *et al.* 1993). For example, determinacy of third order, spikelet meristems requires the activity of the

indeterminate spikelet 1 (*ids1*) gene, which encodes a homolog of *APETALA2* (Chuck, Meeley and Hake 1998).

Recessive mutations in the *ramosa* genes (*Ra1*, *ra2* and *ra3*) affect the binary switch between spikelet pair and branch meristem identity in second order meristems. Of these three mutants, *ra2* and *ra3* have other defects as well, suggesting that their effects on branching may be indirect (our unpublished observations). By contrast, *Ra1* leads to a specific patterning defect without the concomitant loss of any tissue types in the inflorescence.

Ramosa 1 was first described shortly after the rediscovery of genetics in the early part of the 20th century, and was classified as a new subspecies, *Zea ramosa* (Gernert 1912). It was soon recognized as a single gene mutation (Kempton 1921), but the most complete description of the phenotype came when Postlethwait and Nelson first put forth the "switchpoint" concept half a century later (Postlethwait and Nelson 1964). Both the ear and the tassel are many-branched relative to normal, and have a conical appearance. In the tassel, branch length tapers acropetally, while in the ear, branches are most commonly found near the base.

Ears and tassels are many branched due to second order meristems on the main inflorescence axes behaving exclusively as indeterminate branch meristems (Kempton 1921) (Fig. 2). Lower order inflorescence meristems were reported to be unaffected, although *Ra1-ref* ears show poor fertility (Postlethwait and Nelson 1964). Masses of proliferated silks were thought to be responsible for this low fertility by precluding silk exposure to pollen. *Ra1* mutations are also characterized by variable expression of the mutant phenotype, such that either the ear or tassel may be more or less affected in an individual plant (Postlethwait and Nelson 1964).

Thus, the *Ra1* gene product imposes a specific, determinate fate on branches as they arise in the upper portion of the tassel and throughout the ear. In *Ra1* mutants, second order inflorescence meristems in these regions assume branch meristem identity rather than becoming spikelet pairs (Fig. 1).

This identity change could reflect a heterochronic defect, in that late-initiated second order meristems reiterate the fate normally expressed by earlier second order meristems. Alternatively, it could reflect a homeotic effect wherein upper second order meristems assume the fate normally assigned to lower second order meristems. It is also possible that the *Ra1* gene product is a general regulator of determinacy, such that in its absence second order meristems assume a default, indeterminate fate.

Grass inflorescence morphology encompasses a spectacular range of variation (Clifford, 1987; Kellogg and Shaffer 1993), and the maize inflorescence, particularly the ear, is unique among related grasses (Kellogg and Birchler 1993). Indeed, inflorescences of the related panicoid grasses millet and sorghum, as well as of many sedges and rushes, have multiple branches that more closely resemble Ramosa 1 mutants. Potentially, variation in Ramosa 1 expression or regulation may be responsible for these macroevolutionary changes. If so, it may have a critical role in the evolution of graminaceous crop plants.

Interestingly, *Ra1* mutant inflorescences resemble those of related panicoid grasses. Ramosa 1 may account for macroevolutionary change in grass inflorescence architecture. Manipulation of the *ramosa* gene in species such as sorghum and millet would improve yield and could potentially be used to create new crops by altering the structure of the primary inflorescence to resemble rice, wheat and maize.

As can be seen from the foregoing, there is a continuing need in the art for identification of genes and proteins involved in plant architecture and organ development.

It is thus an object of the present invention to provide a novel gene and protein which regulates plant architecture and which may be manipulated to improve health, productivity and yield of plants.

It is yet another object of the invention to provide a DNA sequence of a maize gene the product of which is involved in the regulation plant architecture.

A further object is to provide a mechanism for manipulating meristem identity and to achieve increased yield, to control inflorescence number, branching, arrangement or other reproductive development in plants.

A further object of the present invention is to provide genetic constructs for expression of or inhibition of this gene product, as well as antibodies for recognition of the same.

Finally, it is an object of the present invention to provide genetic material which can used to screen other genomes to identify other genes with similar effects from other plant sources or even from animal sources.

Other objects of the invention will become apparent from the description of the invention which follows.

SUMMARY OF THE INVENTION

According to the invention a novel gene Ramosa 1 (*Ra1*) has been isolated and characterized from maize. This gene encodes a regulatory protein which is intimately involved in the regulation of meristem identity and plant architecture. The *Ra1* gene product is a member of the zinc finger transcription factor family which is characterized by a highly conserved alpha helical motif in the finger region.

The gene encodes a protein product which is intimately involved in the regulation of meristem cell proliferation, particularly in inflorescence development. Ramosa 1 (*Ra1*) is responsible for reducing the number of branches in a plant by promoting the conversion of branch meristems to spikelet pair meristems that produce flowers, or by generally inhibiting the proliferation of branch meristems and thereby allowing spikelet pair meristems to elaborate in their place. Conversely, loss of Ramosa 1 promotes increased branching in both ear and tassel. Therefore, manipulation of the Ramosa 1 gene by mutagenesis or over-expression may be used to 1) alter branch number and 2) alter meristem identity (from indeterminate branch to determinate branch).

Thus the novel gene and protein product of the invention provide a valuable tool for the manipulation of meristem growth and identity, organ development, branching, and flower arrangement, to increase yield, health and stability of plants. Plant architecture has an enormous impact on yield, either by altering the numbers of fruits and seeds produced by the inflorescence (because of extra branches), or by making plants more compact allowing them to be grown under more stringent conditions (e.g., planted at high density or under adverse weather such as heavy rain). It potentially impacts all branches of agriculture including forestry and horticulture. Thus reducing or increasing branching is a key agronomic trait. In maize, increasing the number of branches in the tassel increases pollen yield, which influences overall yield as well as facilitating breeding. In sorghum, or in millets, reducing branch number might result in a maize-like ear and so increase yield that way. Genetic engineering methods known in the art can be used to inhibit expression of the gene or to further induce expression thus controlling the developmental effects regulated thereby, in not only maize but other plants and animals. Further, due to the conserved nature of these zinc finger proteins and of gene function between species, it is expected that other such genes may be identified using the DNA and amino acid sequences herein to characterize other closely related genes from other species.

The invention further comprises novel compositions including protein products and nucleic acid sequences isolated from plants. Also included are expression constructs comprising these sequences as well as transformed cells, vectors and transgenic plants incorporating same. The invention further comprises monoclonal or polyclonal antibodies which recognize the novel proteins of the invention.

The invention also includes methods for manipulating plant architecture, particularly branching, and thus yield of plants by incorporating the expression and or inhibition constructs of the invention. For example, inhibition of expression of this gene via antisense or RNAi would increase branching. This would increase the yield of fruit and seed per plant. In

addition, highly branched tassels would be expected to have increased pollen shed resulting in greater fertility for use in hybrid corn production as well as increased yield.

Increasing Ramosa 1 expression, on the other hand, would be expected to reduce branching. This would be a crucial step in transforming primitive crops such as millet and sorghum into higher yielding derivatives with unbranched maize-like ears, or for designing plants which could be planted at high density.

For purposes of this application the following terms shall have the definitions recited herein. Units, prefixes, and symbols may be denoted in their SI accepted form. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively. Numeric ranges are inclusive of the numbers defining the range and include each integer within the defined range. Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes. Unless otherwise provided for, software, electrical, and electronics terms as used herein are as defined in The New IEEE Standard Dictionary of Electrical and Electronics Terms (5th edition, 1993). The terms defined below are more fully defined by reference to the specification as a whole.

By "amplified" is meant the construction of multiple copies of a nucleic acid sequence or multiple copies complementary to the nucleic acid sequence using at least one of the nucleic acid sequences as a template. Amplification systems include the polymerase chain reaction (PCR) system, ligase chain reaction (LCR) system, nucleic acid sequence based amplification (NASBA, Canteen, Mississauga, Ontario), Q-*Beta* Replicase systems, transcription-based amplification system (TAS), and strand displacement amplification (SDA).

See, e.g., *Diagnostic Molecular Microbiology: Principles and Applications*, D.H. Persing et al., Ed., American Society for Microbiology, Washington, D.C. (1993). The product of amplification is termed an amplicon.

As used herein, "antisense orientation" includes reference to a duplex polynucleotide sequence that is operably linked to a promoter in an orientation where the antisense strand is transcribed. The antisense strand is sufficiently complementary to an endogenous transcription product such that translation of the endogenous transcription product is often inhibited.

As used herein, "chromosomal region" includes reference to a length of a chromosome that may be measured by reference to the linear segment of DNA that it comprises. The chromosomal region can be defined by reference to two unique DNA sequences, i.e., markers.

The term "conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or conservatively modified variants of the amino acid sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations" and represent one species of conservatively modified variation. Every nucleic acid sequence herein that encodes a polypeptide also, by reference to the genetic code, describes every possible silent variation of the nucleic acid. One of ordinary skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine; and UGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide of the present invention is implicit in each described polypeptide sequence and is within the scope of the present invention.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Thus, any number of amino acid residues selected from the group of integers consisting of from 1 to 15 can be so altered. Thus, for example, 1, 2, 3, 4, 5, 7, or 10 alterations can be made. Conservatively modified variants typically provide similar biological activity as the unmodified polypeptide sequence from which they are derived. For example, substrate specificity, enzyme activity, or ligand/receptor binding is generally at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% of the native protein for its native substrate. Conservative substitution tables providing functionally similar amino acids are well known in the art.

The following six groups each contain amino acids that are conservative substitutions for one another:

1) Alanine (A), Serine (S), Threonine (T);
2) Aspartic acid (D), Glutamic acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

See also, Creighton (1984) Proteins W.H. Freeman and Company.

By "encoding" or "encoded", with respect to a specified nucleic acid, is meant comprising the information for translation into the specified protein. A nucleic acid encoding a protein may comprise non-translated sequences (e.g., introns) within translated regions of the nucleic acid, or may lack such intervening non-translated sequences (e.g., as in cDNA). The information by which a protein is encoded is specified by the use of codons. Typically, the amino acid sequence is encoded by the nucleic acid using the "universal" genetic code. However, variants of the universal code, such as are present in

some plant, animal, and fungal mitochondria, the bacterium *Mycoplasma capricolum*, or the ciliate *Macronucleus*, may be used when the nucleic acid is expressed therein.

When the nucleic acid is prepared or altered synthetically, advantage can be taken of known codon preferences of the intended host where the nucleic acid is to be expressed. For example, although nucleic acid sequences of the present invention may be expressed in both monocotyledonous and dicotyledonous plant species, sequences can be modified to account for the specific codon preferences and GC content preferences of monocotyledons or dicotyledons as these preferences have been shown to differ (Murray *et al. Nucl. Acids Res.* 17:477-498 (1989)). Thus, the maize preferred codon for a particular amino acid may be derived from known gene sequences from maize. Maize codon usage for 28 genes from maize plants are listed in Table 4 of Murray *et al., supra.*

As used herein "full-length sequence" in reference to a specified polynucleotide or its encoded protein means having the entire amino acid sequence of, a native (non-synthetic), endogenous, biologically active form of the specified protein. Methods to determine whether a sequence is full-length are well known in the art including such exemplary techniques as northern or western blots, primer extensions, S1 protection, and ribonuclease protection. See, e.g., *Plant Molecular Biology: A Laboratory Manual*, Clark, Ed., Springer-Verlag, Berlin (1997). Comparison to known full-length homologous (orthologous and/or paralogous) sequences can also be used to identify full-length sequences of the present invention. Additionally, consensus sequences typically present at the 5' and 3' untranslated regions of mRNA aid in the identification of a polynucleotide as full-length. For example, the consensus sequence ANNNNAUGG, where the underlined codon represents the N-terminal methionine, aids in determining whether the polynucleotide has a complete 5' end. Consensus sequences at the 3' end, such as polyadenylation sequences, aid in determining whether the polynucleotide has a complete 3' end.

As used herein, "heterologous" in reference to a nucleic acid is a nucleic acid that originates from a foreign species, or, if from the same species, is substantially modified from its native form in composition and/or genomic locus by deliberate human intervention. For example, a promoter operably linked to a heterologous structural gene is from a species different from that from which the structural gene was derived, or, if from the same species, one or both are substantially modified from their original form. A heterologous protein may originate from a foreign species or, if from the same species, is substantially modified from its original form by deliberate human intervention.

By "host cell" is meant a cell which contains a vector and supports the replication and/or expression of the vector. Host cells may be prokaryotic cells such as *E. coli*, or eukaryotic cells such as yeast, insect, amphibian, or mammalian cells. Preferably, host cells are monocotyledonous or dicotyledonous plant cells. A particularly preferred monocotyledonous host cell is a maize host cell.

The term "hybridization complex" includes reference to a duplex nucleic acid structure formed by two single-stranded nucleic acid sequences selectively hybridized with each other.

The term "introduced" in the context of inserting a nucleic acid into a cell, means "transfection" or "transformation" or "transduction" and includes reference to the incorporation of a nucleic acid into a eukaryotic or prokaryotic cell where the nucleic acid may be incorporated into the genome of the cell (e.g., chromosome, plasmid, plastid or mitochondrial DNA), converted into an autonomous replicon, or transiently expressed (e.g., transfected mRNA).

The term "isolated" refers to material, such as a nucleic acid or a protein, which is: (1) substantially or essentially free from components that normally accompany or interact with it as found in its naturally occurring environment. The isolated material optionally comprises material not found with the material in its natural environment; or (2) if the material is in its natural environment, the material has been synthetically (non-naturally)

altered by deliberate human intervention to a composition and/or placed at a location in the cell (e.g., genome or subcellular organelle) not native to a material found in that environment. The alteration to yield the synthetic material can be performed on the material within or removed from its natural state. For example, a naturally occurring nucleic acid becomes an isolated nucleic acid if it is altered, or if it is transcribed from DNA which has been altered, by means of human intervention performed within the cell from which it originates. See, e.g., Compounds and Methods for Site Directed Mutagenesis in Eukaryotic Cells, Kmiec, U.S. Patent No. 5,565,350; *In Vivo* Homologous Sequence Targeting in Eukaryotic Cells; Zarling *et al.*, PCT/US93/03868. Likewise, a naturally occurring nucleic acid (e.g., a promoter) becomes isolated if it is introduced by non-naturally occurring means to a locus of the genome not native to that nucleic acid. Nucleic acids which are "isolated" as defined herein, are also referred to as "heterologous" nucleic acids.

As used herein, "localized within the chromosomal region defined by and including" with respect to particular markers includes reference to a contiguous length of a chromosome delimited by and including the stated markers.

As used herein, "marker" includes reference to a locus on a chromosome that serves to identify a unique position on the chromosome. A "polymorphic marker" includes reference to a marker which appears in multiple forms (alleles) such that different forms of the marker, when they are present in a homologous pair, allow transmission of each of the chromosomes of that pair to be followed. A genotype may be defined by use of one or a plurality of markers.

As used herein, "nucleic acid" or "nucleotide" includes reference to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompasses known analogues having the essential nature of natural nucleotides in that they hybridize to single-stranded nucleic acids in a manner similar to naturally occurring nucleotides (e.g., peptide nucleic acids).

By "nucleic acid library" is meant a collection of isolated DNA or RNA molecules which comprise and substantially represent the entire transcribed fraction of a genome of a specified organism. Construction of exemplary nucleic acid libraries, such as genomic and cDNA libraries, is taught in standard molecular biology references such as Berger and Kimmel, *Guide to Molecular Cloning Techniques, Methods in Enzymology*, Vol. 152, Academic Press, Inc., San Diego, CA (Berger); Sambrook *et al.*, *Molecular Cloning – A Laboratory Manual*, 2nd ed., Vol. 1-3 (1989); and *Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, Eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc. (1994).

As used herein "operably linked" includes reference to a functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates and mediates transcription of the DNA sequence corresponding to the second sequence. Generally, operably linked means that the nucleic acid sequences being linked are contiguous and, where necessary to join two protein coding regions, contiguous and in the same reading frame.

As used herein, the term "plant" can include reference to whole plants, plant parts or organs (e.g., leaves, stems, roots, etc.), plant cells, seeds and progeny of same. Plant cell, as used herein, further includes, without limitation, cells obtained from or found in: seeds, suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen, and microspores. Plant cells can also be understood to include modified cells, such as protoplasts, obtained from the aforementioned tissues. The class of plants which can be used in the methods of the invention is generally as broad as the class of higher plants amenable to transformation techniques, including both monocotyledonous and dicotyledonous plants. As used herein, "polynucleotide" includes reference to a deoxyribopolynucleotide, ribopolynucleotide, or analogs thereof that have the essential nature of a natural ribonucleotide in that they hybridize, under stringent hybridization conditions, to substantially the same nucleotide sequence as naturally occurring nucleotides and/or allow translation into the

same amino acid(s) as the naturally occurring nucleotide(s). A polynucleotide can be full-length or a subsequence of a native or heterologous structural or regulatory gene. Unless otherwise indicated, the term includes reference to the specified sequence as well as the complementary sequence thereof. Thus, DNAs or RNAs with backbones modified for stability or for other reasons as "polynucleotides" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein. It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term polynucleotide as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including among other things, simple and complex cells.

The terms "polypeptide", "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers. The essential nature of such analogues of naturally occurring amino acids is that, when incorporated into a protein, that protein is specifically reactive to antibodies elicited to the same protein but consisting entirely of naturally occurring amino acids. The terms "polypeptide", "peptide" and "protein" are also inclusive of modifications including, but not limited to, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation. It will be appreciated, as is well known and as noted above, that polypeptides are not entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular

polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Further, this invention contemplates the use of both the methionine-containing and the methionine-less amino terminal variants of the protein of the invention.

As used herein "promoter" includes reference to a region of DNA upstream from the start of transcription and involved in recognition and binding of RNA polymerase and other proteins to initiate transcription. A "plant promoter" is a promoter capable of initiating transcription in plant cells whether or not its origin is a plant cell. Exemplary plant promoters include, but are not limited to, those that are obtained from plants, plant viruses, and bacteria which comprise genes expressed in plant cells such as *Agrobacterium* or *Rhizobium*. Examples of promoters under developmental control include promoters that preferentially initiate transcription in certain tissues, such as leaves, roots, or seeds. Such promoters are referred to as "tissue preferred". Promoters which initiate transcription only in certain tissue are referred to as "tissue specific". A "cell type" specific promoter primarily drives expression in certain cell types in one or more organs, for example, vascular cells in roots or leaves. An "inducible" or "repressible" promoter is a promoter which is under environmental control. Examples of environmental conditions that may effect transcription by inducible promoters include anaerobic conditions or the presence of light. Tissue specific, tissue preferred, cell type specific, and inducible promoters constitute the class of "non-constitutive" promoters. A "constitutive" promoter is a promoter which is active under most environmental conditions.

As used herein the term "Ramosa 1" shall include a nucleotide sequence encoding an amino acid sequence having all of the physiological and biological properties of Ramosa 1 as disclosed herein including conservatively modified variants.

As used herein "recombinant" includes reference to a cell or vector, that has been modified by the introduction of a heterologous nucleic acid or that the cell is derived from a cell so modified. Thus, for example, recombinant cells

express genes that are not found in identical form within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under-expressed or not expressed at all as a result of deliberate human intervention. The term "recombinant" as used herein does not encompass the alteration of the cell or vector by naturally occurring events (e.g., spontaneous mutation, natural transformation/transduction/transposition) such as those occurring without deliberate human intervention.

As used herein, a "expression cassette" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements which permit transcription of a particular nucleic acid in a host cell. The recombinant expression cassette can be incorporated into a plasmid, chromosome, mitochondrial DNA, plastid DNA, virus, or nucleic acid fragment. Typically, the recombinant expression cassette portion of an expression vector includes, among other sequences, a nucleic acid to be transcribed, and a promoter.

As used herein the term "*Ral*" or "Ramosa 1" shall include any of the *Ral* amino acid sequences specified herein and their conservatively modified variants which retain the *Ral* biological functions described herein. With respect to a "nucleotide sequence encoding *Ral*" the term includes nucleotide sequences which encode *Ral* and its conservatively modified variants as well as those *Ral* encoding nucleic acid sequences which hybridize under conditions of high stringency to the sequences disclosed herein.

The term "residue" or "amino acid residue" or "amino acid" are used interchangeably herein to refer to an amino acid that is incorporated into a protein, polypeptide, or peptide (collectively "protein"). The amino acid may be a naturally occurring amino acid and, unless otherwise limited, may encompass non-natural analogs of natural amino acids that can function in a similar manner as naturally occurring amino acids.

The term "selectively hybridizes" includes reference to hybridization, under stringent hybridization conditions, of a nucleic acid sequence to a

specified nucleic acid target sequence to a detectable greater degree (e.g., at least 2-fold over background) than its hybridization to non-target nucleic acid sequences and to the substantial exclusion of non-target nucleic acids. Selectively hybridizing sequences typically have about at least 80% sequence identity, preferably 90% sequence identity, and most preferably 100% sequence identity (i.e., complementary) with each other.

The term "stringent conditions" or "stringent hybridization conditions" includes reference to conditions under which a probe will hybridize to its target sequence, to a detectable greater degree than to other sequences (e.g., at least 2-fold over background). Stringent conditions are sequence-dependent and may be different in different circumstances. By controlling the stringency of the hybridization and/or washing conditions, target sequences can be identified which are 100% complementary to the probe (homologous probing). Alternatively, stringency conditions can be adjusted to allow some mismatching in sequences so that lower degrees of similarity are detected (heterologous probing). Generally, a probe is less than about 1000 nucleotides in length, optionally less than 500 nucleotides in length.

Typically, stringent conditions will be those in which the salt concentration is less than about 1.5 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (*e.g.*, 10 to 50 nucleotides) and at least about 60°C for long probes (*e.g.*, greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulphate) at 37°C, and a wash in 1X to 2X SSC (20X SSC = 3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.5X to 1X SSC at 55 to 50°C. Exemplary high stringency conditions include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1X SSC at 60 to 65°C.

Specificity is typically the function of post-hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the T_m can be approximated from the equation of Meinkoth and Wahl, *Anal. Biochem.*, 138:267-284 (1984): T_m=81.5°C + 16.6 (log M) + 0.41 (%GC) –0.61 (% form) – 500/L; where M is the molarity of monovalent cations, %GC is the percentage of guanosine and cytosine nucleotides in the DNA, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the complementary target sequence hybridizes to a perfectly matched probe. T_m is reduced by about 1°C for each 1% of mismatching; thus, T_m, hybridization and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with ≥90% identity are sought, the T_m can be decreased 10°C. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3, or 4°C lower than the thermal melting point (T_m); moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or 10°C lower than the thermal melting point (T_m); low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15, or 20°C lower than the thermal melting point (T_m). Using the equation, hybridization and wash compositions, and desired T_m, those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results in a T_m of less than 45°C (aqueous solution) or 32°C (formamide solution) it is preferred to increase the SSC concentration so that a higher temperature can be used. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Laboratory Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Acids Probes*, Part I, Chapter 2, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995).

As used herein, the term "structural gene" includes any nucleotide sequence the expression of which is desired in a plant cell. A structural gene can include an entire sequence encoding a protein, or any portion thereof. Examples of structural genes are included hereinafter are intended for illustration and not limitation.

As used herein, "transgenic plant" includes reference to a plant which comprises within its genome a heterologous polynucleotide. Generally, the heterologous polynucleotide is stably integrated within the genome such that the polynucleotide is passed on to successive generations. The heterologous polynucleotide may be integrated into the genome alone or as part of a recombinant expression cassette. "Transgenic" is used herein to include any cell, cell line, callus, tissue, plant part or plant, the genotype of which has been altered by the presence of heterologous nucleic acid including those transgenics initially so altered as well as those created by sexual crosses or asexual propagation from the initial transgenic. The term "transgenic" as used herein does not encompass the alteration of the genome (chromosomal or extra-chromosomal) by conventional plant breeding methods or by naturally occurring events such as random cross-fertilization, non-recombinant viral infection, non-recombinant bacterial transformation, non-recombinant transposition, or spontaneous mutation.

As used herein, "vector" includes reference to a nucleic acid used in transfection of a host cell and into which can be inserted a polynucleotide. Vectors are often replicons. Expression vectors permit transcription of a nucleic acid inserted therein.

The following terms are used to describe the sequence relationships between two or more nucleic acids or polynucleotides: (a) "reference sequence", (b) "comparison window", (c) "sequence identity", (d) "percentage of sequence identity", and (e) "substantial identity".

(a) As used herein, "reference sequence" is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset or the

entirety of a specified sequence; for example, as a segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence.

(b) As used herein, "comparison window" includes reference to a contiguous and specified segment of a polynucleotide sequence, wherein the polynucleotide sequence may be compared to a reference sequence and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides in length, and optionally can be 30, 40, 50, 100, or longer. Those of skill in the art understand that to avoid a high similarity to a reference sequence due to inclusion of gaps in the polynucleotide sequence, a gap penalty is typically introduced and is subtracted from the number of matches.

Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman, *Adv. Appl. Math.* 2:482 (1981); by the homology alignment algorithm of Needleman and Wunsch, *J. Mol. Biol.* 48:443 (1970); by the search for similarity method of Pearson and Lipman, *Proc. Natl. Acad. Sci.* 85:2444 (1988); by computerized implementations of these algorithms, including, but not limited to: CLUSTAL in the PC/Gene program by Intelligenetics, Mountain View, California; GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, Wisconsin, USA; the CLUSTAL program is well described by Higgins and Sharp, *Gene* 73:237-244 (1988); Higgins and Sharp, CABIOS 5:151-153 (1989); Corpet, et al., *Nucleic Acids Research* 16:10881-90 (1988); Huang, *et al.*, *Computer Applications in the Biosciences* 8:155-65 (1992), and Pearson, *et al.*, *Methods in Molecular Biology* 24:307-331 (1994). The BLAST family of programs which can be used for database similarity searches includes: BLASTN for nucleotide query sequences against nucleotide database sequences; BLASTX for nucleotide query sequences against protein database

sequences; BLASTP for protein query sequences against protein database sequences; TBLASTN for protein query sequences against nucleotide database sequences; and TBLASTX for nucleotide query sequences against nucleotide database sequences. See, *Current Protocols in Molecular Biology*, Chapter 19, Ausubel, et al., Eds., Greene Publishing and Wiley-Interscience, New York (1995).

Unless otherwise stated, sequence identity/similarity values provided herein refer to the value obtained using the BLAST 2.0 suite of programs using default parameters. Altschul et a., *Nucleic Acids Res.* 25:3389-3402 (1997). Software for performing BLAST analyses is publicly available, e.g., through the National Center for Biotechnology-Information (http://www.hcbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both

strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915).

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul, *Proc. Natl. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance.

BLAST searches assume that proteins can be modeled as random sequences. However, many real proteins comprise regions of nonrandom sequences which may be homopolymeric tracts, short-period repeats, or regions enriched in one or more amino acids. Such low-complexity regions may be aligned between unrelated proteins even though other regions of the protein are entirely dissimilar. A number of low-complexity filter programs can be employed to reduce such low-complexity alignments. For example, the SEG (Wooten and Federhen, *Comput. Chem.*, 17:149-163 (1993)) and XNU (Claverie and States, *Comput. Chem.*, 17:191-201 (1993)) low-complexity filters can be employed alone or in combination.

(c) As used herein, "sequence identity" or "identity" in the context of two nucleic acid or polypeptide sequences includes reference to the residues in the two sequences which are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g. charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences

which differ by such conservative substitutions are said to have "sequence similarity" or "similarity". Means for making this adjustment are well-known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, e.g., according to the algorithm of Meyers and Miller, *Computer Applic. Biol. Sci.*, 4:11-17 (1988) e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California, USA).

(d) As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

(e)(I) The term "substantial identity" of polynucleotide sequences means that a polynucleotide comprises a sequence that has at least 70% sequence identity, preferably at least 80%, more preferably at least 90% and most preferably at least 95%, compared to a reference sequence using one of the alignment programs described using standard parameters. One of skill will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning and the like. Substantial identity of amino acid sequences for

these purposes normally means sequence identity of at least 60%, or preferably at least 70%, 80%, 90%, and most preferably at least 95%.

Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other under stringent conditions. However, nucleic acids which do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This may occur, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. One indication that two nucleic acid sequences are substantially identical is that the polypeptide which the first nucleic acid encodes is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

(e)(ii) The terms "substantial Identity" in the context of a peptide indicates that a peptide comprises a sequence with at least 70% sequence identity to a reference sequence, preferably 80%, or preferably 85%, most preferably at least 90% or 95% sequence identity to the reference sequence over a specified comparison window. Optionally, optimal alignment is conducted using the homology alignment algorithm of Needleman and Wunsch, *J. Mol. Biol.* 48:443 (1970). an indication that two peptide sequences are substantially identical is that one peptide is immunologically reactive with antibodies raised against the second peptide. Thus, a peptide is substantially identical to a second peptide, for example, where the two peptides differ only by a conservative substitution. Peptides which are "substantially similar" share sequences as noted above except that residue positions which are not identical may differ by conservative amino acid changes.

DESCRIPTION OF THE FIGURES

Figure 1 is a schematic of normal and *Ra1* mutant inflorescence development. Meristem types (I.M., etc.) are defined in text.

Figure 2 is an image depicting normal and *Ra1* mutant inflorescences. (A) Mature, normal tassel. (B) Mature, *Ra1-ref* tassel. (C) An immature, highly

branched *Ra1-ref* ear. (D) A range of mutant, mature ear phenotypes. The two ears on the left, from a *Ra1-m* allele, are nearly normal. The highly branched ear on the right will be partially sterile.

Figure 3 depicts an *Ra1* allelic series. Mutants have been converged at least 3 times into B73. Tassels are at anthesis. (A) A normal, B73 tassel. (B) The weak allele *Ra1-IHO*. (C) *Ra1-RS*, an intermediate strength allele. *Ra1-63.3359* show a similar phenotype. (D) *Ra1-Mum1*, one of several strong alleles.

Figure 4 depicts SEMs of immature ears. (A) Normal. Note smooth surface (arrow) opposite carpel . Each glume (g) subtends a single spikelet. (B) *Ra1-ref*. Note additional organs (arrows) initiating opposite each carpel, which is also rotated. Glume (g) subtends multiple spikelets. (C) shows a normal flower at later stage, as (B); no primordia initiate opposite carpel.

Figure 5 depicts *Ra1-m* alleles recovered from the *Spm* screen. (A) This *Ra1-m2* revertant sector encompasses approximately 75% of the circumference of the tassel. (B) A *Ra1-m3* half-tassel sector head-on. Spikelets in the revertant sector are appressed to the central spike, with branches on the opposite side. (C) A small *Ra1-m2* revertant sector, one row of spikelets wide, is present on the main axis adjacent to the red line.

Figure 6 shows *Spm*-containing restriction enzyme fragments co-segregate with the *Ra1-m2* and *Ra1-m3* mutations. Our genetic analysis (see text) demonstrates co-segregation with many more individuals than shown here. (A) Genomic DNA cut with *Hind* III. Arrows indicate 16 kb (upper left) and 3.5 kb (lower right) fragments. (B) Genomic DNA cut with *Eco*RI. Arrow indicates the 5 kb fragment.

Figure 7 depicts the mutations of *Ra1* gene.

Figure 8 shows the homology and alignment of *Ra1* with Arapidopsis zinc finger proteins.

Figure 9 shows the cDNA and amino acid sequence of the maize *Ra1* gene, SEQ ID NO:1.

Figure 10 depicts *Ra1* expressed in discrete domains within the developing inflorescence. (A) SEM of immature tassel inflorescence. (B-E) *In situ* hybridization analysis of *Ra1* expression. (B) Full-length *Ra1* cDNA shows general staining in inflorescence, branch, spikelet pair, and spikelet meristems, and more intense, punctate staining in the axils of spikelet pair and spikelet meristems. (C) A shorter probe from the 5' end of the gene specifically detects expression in a broad axillary domain of spikelet pair and young spikelet meristems. (D) At later development stages, *Ra1* is expressed on the abaxial side of the pedicel of maturing spikelets. (E) Closep-up of abaxial expression in spikelet pedicels as in D.

Figure 11 shows the *Ra1* genomic DNA sequence from B73, 4936 bp, annotated with insertion mutations. The laboratory names for the mutant alleles are indicated above the sequence. The identity of the insertion follows the allele designation, preceded by 2 colons (::). Transoposable element insertion sites are indicated by a double underline of the 3 bp target site duplication. "CACTA element" refers to an unclassified transposable element of the CACTA class. The transcription start site is indicated above the sequence by an arrow (➔) (SEQ ID NO:6).

Figure 12 shows the *Ra1-R* (also termed herein *Ra1-ref*) mutant allele cDNA sequence (SEQ ID NO:4).

Figure 13 shows the genomic DNA sequence, *ramosa1* ortholog from sugar cane (Saccharum oficinarum). The sugar cane sequence encodes an amino acid sequence that contains the same novel sequence in the zinc finger region as is present in the *ra1* mutant reference allele (QGLEGN), suggesting that the sugar cane ortholog may be nonfunctional or reduced function, consistent with the fully branched phenotype of the sugar cane inflorescence (SEQ ID NO:3).

Figure 14 shows the Genomic DNA sequence, *ramosa1* ortholog from teosinte (*Zea mays parviglumis*) (SEQ ID NO:5).

DETAILED DESCRIPTION OF THE INVENTION

Applicants have discovered a regulatory meristem development gene isolated from maize that is involved in plant architecture. The gene is likely a member of the zinc finger protein family which has been unexplored in maize. Mutations in the *Ra1* gene product as shown herein cause indeterminate second order meristems, resulting in highly branched inflorescence both in the ear and the tassel. The *Ra1* gene and protein product can regulate branching both negatively and positively depending upon the state of the gene. The wild type protein acts to reduce the number of branches in a plant by promoting conversion of branch meristems to spikelet pair meristems that produce flowers while the mutant forms or loss of *Ra1* promotes increased branching in the ear and tassel.

Thus in one embodiment of the invention, the *Ra1* gene or its protein product can be used in regulation of meristem development and branching, both if overexpressed or if the activity is suppressed by mutation or by other mechanisms such as for instance by antisense expression, homologous recombination or co-suppression mechanisms.

According to the invention the *Ra1* gene from maize has been cloned using Spm transposon mutagenesis and sequenced. The gene has been shown to be expressed in the expected spikelet pair meristems by in situ hybridization. According to the invention a number of mutant alleles have been characterized. Interestingly, the normal allele, found in maize inbreds, has a conservative amino-acid change in the presumed DNA binding domain, which may result in weak branching in the tassel of maize inbred lines.

The invention herein in its broadest sense contemplates the discovery of the existence of an *Ra1* gene in plants that is associated among other things with meristem development, inflorescence development, and plant architecture. The discovery of the existence of this type of gene creates numerous opportunities for manipulation of inflorescence development, and/or branching in plants in general. Due to the highly conserved nature of the gene product, and the highly conserved nature of inflorescence development

within the phylogentically broad group of organisms comprising the grass family, it is expected that this gene or ones substantially equivalent thereto may be identified from other plants with similar meristem specific functions. These homologs are intended to be within the scope of this invention and have been identified in sugar cane (Figure 13) and teosinte (Figure 14). Similarly, the protein product disclosed here also may be used for other plants and many other mutants may be either engineered by those of skill in the art or isolated from other species. Homologous proteins or mutants as described herein and as isolated from other plants are also intended to be within the scope of this invention. It is likely that this gene controls the differences observed among grasses in inflorescence architecture. We have isolated homologs from a variety of related grasses and we are sequencing them. We have sequenced related genes from sugar cane and teosinte. Manipulation of this gene in primitive grasses may allow them to be used as crops.

This invention further contemplates methods of controlling organ development, cell proliferation, flower development etc by manipulating *Ra1* genes in plants through genetic engineering techniques which are known and commonly used by those of skill in the art. Such methods include but are in no way limited to generation of increased seed number, flower or organ number, arrangement, size, etc., as well as other tissue specific regulation based upon expression of the gene at time, spatial and developmental periods.

In yet another embodiment the invention comprises regulatory sequences associated with the novel Ra1 gene of the invention. This regulatory region may be used to achieve expression of heterologous genes in spikelet pair meristems or other tissues associated with and during periods of plant architectural development.

The invention in one aspect comprises expression constructs comprising a DNA sequence which encodes upon expression a *Ra1* gene product operably linked to a promoter to direct expression of the protein. These constructs are then introduced into plant cells using standard molecular biology techniques.

The invention can be also be used for hybrid plant or seed production, once transgenic inbred parental lines have been established.

In another aspect the invention involves the inhibition of an *Ral* gene product in plants through introduction of a construct designed to inhibit the same gene product. The design and introduction of such constructs based upon known DNA sequences is known in the art and includes such technologies as antisense RNA or DNA, co-suppression or any other such mechanism. Several of these mechanisms are described and disclosed in United States Patent 5,686,649 to Chua et. al, which is hereby expressly incorporated herein by reference.

The methods of the invention described herein may be applicable to any species of plant.

The polynucleotides useful in the invention can be formed from a variety of different polynucleotides (e.g., genomic or cDNA, RNA, synthetic oligonucleotides, and polynucleotides), as well as by a variety of different techniques. As used herein, a polynucleotide is a sequence of either eukaryotic or prokaryotic synthetic invention.

The nucleotide constructs of the present invention will share similar elements, which are well known in the art of plant molecular biology. For example, in each construct the DNA sequences of interest will preferably be operably linked (i.e., positioned to ensure the functioning of) to a promoter which allows the DNA to be transcribed (into an RNA transcript) and will comprise a vector which includes a replication system. In preferred embodiments, the DNA sequence of interest will be of exogenous origin in an effort to prevent co-suppression of the endogenous genes.

Promoters (and other regulatory elements) may be heterologous (i.e., not naturally operably linked to a DNA sequence from the same organism). Promoters useful for expression in plants are known in the art and can be inducible, constitutive, tissue-specific, derived from eukaryotes, prokaryotes or viruses, or have various combinations of these characteristics.

In choosing a promoter to use in the methods of the invention, it may be desirable to use a tissue-specific or developmentally regulated promoter. A tissue-specific or developmentally regulated promoter is a DNA sequence which regulates the expression of a DNA sequence selectively in the cells/tissues of a plant critical to seed set and/or function and/or limits the expression of such a DNA sequence to the period of seed maturation in the plant. Any identifiable promoter may be used in the methods of the present invention which causes the desired temporal and spatial expression.

Promoters which are timed to stress include the following: barley promoter B22E: 69 NAL Call No. 442.8 Z34 "Primary Structure of a Novel Barley Gene Differentially Expressed in Immature Alleurone Layers," Klemsdae, S.S. et al., Springer Int'l 1991 Aug., Molecular and General Genetics, Vol. 228(1/2) p. 9-16, 1991. Expression of B22E is specific to the pedicel in developing maize kernels, Zag2: 134 NAL Call. No.: QK725.P532 Identification and molecular characterization of ZAG1, the maize homolog of the Arabidopsis floral homeotic gene AGAMOUS. Schmidt, R.J.; Veit, B.; Mandel, M.A.; Mena, M.; Hake, S.; Yanofsky, M.F. Rockville, MD: American Society of Plant Physiologists, c1989-; 1993 Jul. The Plant Cell v. 5(7): p 729-737; 1993 Jul. includes references. Zag2 transcripts can be detected 5 days prior to pollination to 7 to 8 DAP, and directs expression in the carpel of developing female inflorescences and Cim1 which is specific to the nucleus of developing maize kernels. Cim1 transcript is detected 4 to 5 days before pollination to 6 to 8 DAP. Other useful promoters include any promoter which can be derived from a gene whose expression is maternally associated with developing female florets.

Other promoters which are seed or embryo specific and may be useful in the invention include patatin (potato tubers) (Rocha-Sosa, M., et al. (1989) EMBO J. 8:23-29), convicilin, vicilin, and legumin (pea cotyledons) (Rerie, W.G., et al. (1991) Mol. Gen. Genet. 259:149-157; Newbigin, E.J., et al. (1990) Planta 180:461-470; Higgins, T.J.V., et al. (1988) Plant. Mol. Biol. 11:683-695), zein (maize endosperm) (Schemthaner, J.P., et al. (1988) EMBO J. 7:1249-

1255), phaseolin (bean cotyledon) (Segupta-Gopalan, C., et al. (1985) Proc. Natl. Acad. Sci. U.S.A. 82:3320-3324), phytohemagglutinin (bean cotyledon) (Voelker, T. et al. (1987) EMBO J. 6:3571-3577), B-conglycinin and glycinin (soybean cotyledon) (Chen, Z-L, et al. (1988) EMBO J. 7:297-302), glutelin (rice endosperm), hordein (barley endosperm) (Marris, C., et al. (1988) Plant Mol. Biol. 10:359-366), glutenin and gliadin (wheat endosperm) (Colot, V., et al. (1987) EMBO J. 6:3559-3564), and sporamin (sweet potato tuberous root) (Hattori, T., et al. (1990) Plant Mol. Biol. 14:595-604). Promoters of seed-specific genes operably linked to heterologous coding regions in chimeric gene constructions maintain their temporal and spatial expression pattern in transgenic plants. Such examples include *Arabidopsis thaliana* 2S seed storage protein gene promoter to express enkephalin peptides in *Arabidopsis* and *Brassica napus* seeds (Vanderkerckhove et al., Bio/Technology 7:L929-932 (1989)), been lectin and bean β-phaseolin promoters to express luciferase (Riggs et al., Plant Sci. 63:47-57 (1989)), and wheat glutenin promoters to express chloramphenicol acetyl transferase (Colot et al., EMBO J 6:3559-3564 (1987)).

Any inducible promoter can be used in the instant invention. *See* Ward *et al. Plant Mol. Biol.*22: 361-366 (1993). Exemplary inducible promoters include, but are not limited to, that from the ACEl system which responds to copper (Mett *et al. PNAS* 90: 4567-4571 (1993)); In2 gene from maize which responds to benzenesulfonamide herbicide safeners (Hershey *et al., Mol. Gen. Genetics* 227: 229-237 (1991) and Gatz *et al., Mol. Gen. Genetics* 243: 32-38 (1994)) or Tet repressor from Tn10 (Gatz *et al., Mol. Gen. Genet.* 227: 229-237 (1991). A particularly preferred inducible promoter is a promoter that responds to an inducing agent to which plants do not normally respond. An exemplary inducible promoter is the inducible promoter from a steroid hormone gene, the transcriptional activity of which is induced by a glucocorticosteroid hormone. Schena *et al., Proc. Natl. Acad. Sci. U.S.A.* 88: 0421 (1991).

Many different constitutive promoters can be utilized in the instant invention. Exemplary constitutive promoters include, but are not limited to, the promoters from plant viruses such as the 35S promoter from CaMV (Odell *et al., Nature* 313: 810-812 (1985) and the promoters from such genes as rice actin (McElroy et al., Plant Cell 2: 163-171 (1990)); ubiquitin (Christensen et al., Plant Mol. Biol 12: 619-632 (1989) and Christensen et al., Plant Mol. Biol. 18: 675-689 (1992)): pEMU (Last et al., Theor. Appl. Genet. 81: 581-588 (1991)); MAS (Velten et al., EMBO J. 3: 2723-2730 (1984)) and maize H3 histone (Lepetit et al., Mol. Gen. Genet. 231: 276-285 (1992) and Atanassova et al., Plant Journal 2 (3): 291-300 (1992)).

The ALS promoter, a Xbal/*Nco*l fragment 5' to the *Brassica napus* ALS3 structural gene (or a nucleotide sequence that has substantial sequence similarity to said *Xbal/Nco*l fragment), represents a particularly useful constitutive promoter. See PCT application WO96/30530.

Transport of protein produced by transgenes to a subcellular compartment such as the chloroplast, vacuole, peroxisome, glyoxysome, cell wall or mitochondrion, or for secretion into the apoplast, is accomplished by means of operably linking the nucleotide sequence encoding a signal sequence to the 5' and/or 3' region of a gene encoding the protein of interest. Targeting sequences at the 5' and/or 3' end of the structural gene may determine, during protein synthesis and processing, where the encoded protein is ultimately compartmentalized. The presence of a signal sequence directs a polypeptide to either an intracellular organelle or subcellular compartment or for secretion to the apoplast. Many signal sequences are known in the art. See, for example, Sullivan, T., "Analysis of Maize *Brittle*-1 Alleles and a Defective Suppressor-Mutator-Induced Mutable Allele", The Plant Cell, 3:1337-1348 (1991), Becker *et al., Plant Mol. Biol.*20: 49 (1992), Close, P.S., Master's Thesis, Iowa State University (1993), Knox, C., *et al.,* "Structure and Organization of Two Divergent Alpha-Amylase Genes From Barley", *Plant Mol.Biol.* 9: 3-17 (1987), Lerner *et al., Plant Physiol.*91: 124-129 (1989), Fontes *et al.,Plant Cell* 3: 483-496 (1991), Matsuoka *et al., Proc. Natl. Acad. Sci.* 88: 834 (1991), Gould et al.,

J. Cell Biol 108: 1657 (1989), Creissen *et al., Plant J.* 2: 129 (1991), Kalderon, D., Robers, B., Richardson, W., and Smith A., "A short amino acid sequence able to specify nuclear location", *Cell* 39: 499-509 (1984), Stiefel, V., Ruiz-Avila, L., Raz R., Valles M., Gomez J., Pages M., Martinez-Izquierdo J., Ludevid M., Landale J., Nelson T., and Puigdomenech P., "Expression of a maize cell wall hydroxyproline-rich glycoprotein gene in early leaf and root vascular differentiation", *Plant Cell* 2: 785-793 (1990).

Selection of an appropriate vector is relatively simple, as the constraints are minimal. The minimal traits of the vector are that the desired nucleic acid sequence be introduced in a relatively intact state. Thus, any vector which will produce a plant carrying the introduced DNA sequence should be sufficient. Typically, an expression vector contains (1) prokaryotic DNA elements encoding for a bacterial replication origin and an antibiotic resistance marker to provide for the growth and selection of the expression vector in a bacterial host; (2) DNA elements that control initiation of transcription, such as a promoter; (3) DNA elements that control the processing of transcripts such as transcription termination/polyadenylation sequences; and (4) a reporter gene. Useful reporter genes include β-glucuronidase, β-galactosidase, chloramphynical acetyltransferase, luciferase, kanamycin or the herbicide resistance genes PAT and BAR. Preferably, the reporter gene is kanamyacin or the herbicide resistance genes PAT and BAR. The BAR or PAT gene is used with the selecting agent Bialaphos, and is used as a preferred selection marker gene for plant transformation (Spencer, et al. (1990) J. Thero. Appl'd Genetics 79:625-631).

One commonly used selectable marker gene for plant transformation is the neomycin phosphotransferase II *(nptII)* gene, isolated from transposon Tn5, which when placed under the control of plant regulatory signals confers resistance to kanamycin. Fraley *et al., Proc. Natl. Acad. Sci. U.S.A.,* 80: 4803 (1983). Another commonly used selectable marker gene is the hygromycin phosphotransferase gene which confers resistance to the antibiotic hygromycin. Vanden Elzen et al., Plant Mol. Biol., 5: 299 (1985).

Additional selectable marker genes of bacterial origin that confer resistance to antibiotics include gentamycin acetyl transferase, streptomycin phosphotransferase, aminoglycoside- 3′ -adenyl transferase, the bleomycin resistance determinant. Hayford *et al.*, Plant Physiol. 86: 1216 (1988), Jones et al., Mol. Gen. Genet., 210: 86 (1987), Svab et al., Plant Mol.. Biol.. 14: 197 (1990), Hille et al., Plant Mol. Biol. 7: 171 (1986). Other selectable marker genes confer resistance to herbicides such as glyphosate, glufosinate or broxynil. Comai et al., Nature 317: 741-744 (1985), Gordon-Kamm et al., Plant Cell 2: 603-618 (1990) and Stalker et al., Science 242: 419-423 (1988).

Other selectable marker genes for plant transformation are not of bacterial origin. These genes include, for example, mouse dihydrofolate reductase, plant 5 - *enol*pyruvylshikimate-3 -phosphate synthase and plant acetolactate synthase. Eichholtz et al., Somatic Cell Mol. Genet. 13: 67 (1987), Shah et al., Science 233: 478 (1986), Charest et al., Plant Cell Rep. 8: 643 (1990).

Another class of marker genes for plant transformation require screening of presumptively transformed plant cells rather than direct genetic selection of transformed cells for resistance to a toxic substance such as an antibiotic. These genes are particularly useful to quantify or visualize the spatial pattern of expression of a gene in specific tissues and are frequently referred to as reporter genes because they can be fused to a gene or gene regulatory sequence for the investigation of gene expression. Commonly used genes for screening presumptively transformed cells include β-glucuronidase *(GUS)*, β-galactosidase, luciferase and chloramphenicol acetyltransferase. Jefferson, R.A., Plant Mol. Biol. Rep. 5: 387 (1987)., Teeri et al., EMBO J. 8: 343 (1989), Koncz et al., Proc. Natl. Acad. Sci. U.S.A. 84:131 (1987), De Block et al., EMBO J. 3: 1681 (1984). Another approach to the identification of relatively rare transformation events has been use of a gene that encodes a dominant constitutive regulator of the Zea mays anthocyanin pigmentation pathway. Ludwig et al., Science 247: 449 (1990).

Recently, *in vivo* methods for visualizing GUS activity that do not require destruction of plant tissue have been made available. Molecular Probes Publication 2908, Imagene Green, p. 1-4 (1993) and Naleway *et al., J. Cell Biol.* 115: 15la (1991). However, these *in vivo* methods for visualizing GUS activity have not proven useful for recovery of transformed cells because of low sensitivity, high fluorescent backgrounds, and limitations associated with the use of luciferase genes as selectable markers.

More recently, a gene encoding Green Fluorescent Protein (GFP) has been utilized as a marker for gene expression in prokaryotic and eukaryotic cells. Chalfie *et al., Science* 263: 802 (1994). GFP and mutants of GFP may be used as screenable markers.

Genes included in expression vectors must be driven by a nucleotide sequence comprising a regulatory element, for example, a promoter. Several types of promoters are now well known in the transformation arts, as are other regulatory elements that can be used alone or in combination with promoters.

A general description of plant expression vectors and reporter genes can be found in Gruber, et al. (Gruber et al. (1993) Vectors for Plant Transformation. In: Methods in Plant Molecular Biology and Biotechnology. Glich et al., eds. (CRC Press), pp. 89-119.

Expression vectors containing genomic or synthetic fragments can be introduced into protoplast or into intact tissues or isolated cells. Preferably, expression vectors are introduced into intact tissue. General methods of culturing plant tissues are provided for example by Maki, et al. (Maki, et al. (1993) Procedures for Introducing Foreign DNA into Plants: In: Methods in Plant Molecular Biology & Biotechnology; Glich et al. eds. (CRC Press), pp. 67-88; Philips, et al. (1988) Cell-Tissue Culture and In Vitro Manipulation. In Corn & Corn Improvement, 3rd ed. Sprague, et al. eds. (American Society of Agronomy Inc.), pp. 345-387).

Methods of introducing expression vectors into plant tissue include the direct transfection or co-cultivation of plant cell with *Agrobacterium tumefaciens* (Horsch et al. (1985) Science, 227:1229). Descriptions of

Agrobacterium vector systems and methods for *Agrobacterium*-mediated gene transfer are provided by Gruber et al. (supra).

Numerous methods for plant transformation have been developed, including biological and physical, plant transformation protocols. See, for example, Miki *et al.*, "Procedures for Introducing Foreign DNA into Plants" in *Methods in Plant Molecular Biology and Biotechnology*, Glick, B.R. and Thompson, J.E. Eds. (CRC Press, Inc., Boca Raton, 1993) pages 67-88. In addition, expression vectors and *in vitro* culture methods for plant cell or tissue transformation and regeneration of plants are available. See, for example, Gruber *et al.*, "Vectors for Plant Transformation" in *Methods in Plant Molecular Biology and Biotechnology*, Glick, B.R. and Thompson, J.E. Eds. (CRC Press, Inc., Boca Raton, 1993) pages 89-119.

Agrobacterium-mediated Transformation

One method for introducing an expression vector into plants is based on the natural transformation system of *Agrobacterium*. See, for example, Horsch *et al.*, *Science* 227: 1229 (1985). *A. tumefaciens* and *A. rhizogenes* are plant pathogenic soil bacteria which genetically transform plant cells. The Ti and Ri plasmids of *A. tumefaciens* and *A. rhizogenes*, respectively, carry genes responsible for genetic transformation of the plant. See, for example, Kado, C.I., Crit. Rev. Plant. Sci. 10: 1 (1991). Descriptions of *Agrobacterium* vector systems and methods for *Agrobacterium*-mediated gene transfer are provided by Gruber et al., supra, Miki et al., supra, and Moloney et al., Plant Cell Reports 8: 238 (1989). See also, U.S. Patent No. 5,591,616, issued Jan. 7, 1997.

Direct Gene Transfer

Despite the fact the host range for *Agrobacterium*-mediated transformation is broad, some major cereal crop species and gymnosperms have generally been recalcitrant to this mode of gene transfer, even though some success has recently been achieved in rice and maize. Hiei et al., The

Plant Journal 6: 271-282 (1994); U.S. Patent No. 5,591,616, issued Jan. 7, 1997. Several methods of plant transformation, collectively referred to as direct gene transfer, have been developed as an alternative to Agrobacterium-mediated transformation.

A generally applicable method of plant transformation is microprojectile-mediated transformation wherein DNA is carried on the surface of microprojectiles measuring 1 to 4 mm. The expression vector is introduced into plant tissues with a biolistic device that accelerates the microprojectiles to speeds of 300 to 600 m/s which is sufficient to penetrate plant cell walls and membranes. Sanford et al., Part. Sci. Technol. 5: 27 (1987), Sanford, J.C., Trends Biotech. 6: 299 (1988), Klein et al., Bio/Technology 6: 559-563 (1988), Sanford, J.C., Physiol Plant 79: 206 (1990), Klein et al., Biotechnology 10: 268 (1992). In maize, several target tissues can be bombarded with DNA-coated microprojectiles in order to produce transgenic plants, including, for example, callus (Type I or Type II), immature embryos, and meristematic tissue.

Another method for physical delivery of DNA to plants is sonication of target cells. Zhang et al., Bio/Technology 9: 996 (1991). Alternatively, liposome or spheroplast fusion have been used to introduce expression vectors into plants. Deshayes et al., EMBO J., 4: 2731 (1985), Christou et al., Proc Natl. Acad. Sci. U.S.A. 84: 3962 (1987). Direct uptake of DNA into protoplasts using CaCl2 precipitation, polyvinyl alcohol or poly-L-ornithine have also been reported. Hain et al., Mol. Gen. Genet.199: 161 (1985) and Draper et al., Plant Cell Physiol.23: 451 (1982). Electroporation of protoplasts and whole cells and tissues have also been described. Donn et al., In Abstracts of VIIth International Congress on Plant Cell and Tissue Culture IAPTC, A2-38, p 53 (1990); D'Halluin et al., Plant Cell 4: 1495-1505 (1992) and Spencer et al., Plant Mol. Biol. 24: 51-61 (1994).

Following transformation of target tissues, expression of the above-described selectable marker genes allows for preferential selection of

transformed cells, tissues and/or plants, using regeneration and selection methods now well known in the art.

After transformation of a plant cell or plant, plant cells or plants transformed with the desired DNA sequences integrated into the genome can be selected by appropriate phenotypic markers. Phenotypic markers are known in the art and may be used in this invention.

Confirmation of transgenic plants will typically be based on an assay or assays or by simply measuring stress response. Transformed plants can be screened by biochemical, molecular biological, and other assays. Various assays may be used to determine whether a particular plant, plant part, or a transformed cell shows an increase in enzyme activity or carbohydrate content. Typically, the change in expression or activity of a transformed plant will be compared to levels found in wild type (e.g., untransformed) plants of the same type. Preferably, the effect of the introduced construct on the level of expression or activity of the endogenous gene will be established from a comparison of sibling plants with and without the construct. Protein levels can be measured, for example, by Northern blotting, primer extension, quantitative or semi-quantitative PCR (polymerase chain reaction), and other methods well known in the art (See, e.g., Sambrook, et al. (1989). Molecular Cloning, A Laboratory Manual, second edition (Cold Spring Harbor Laboratory Press), Vols. 1-3). Protein can be measured in a number of ways including immunological methods (e.g., by Elisa or Western blotting). Protein activity can be measured in various assays as described in Smith (Smith, A.M. (1990). In: Methods in Plant Biochemistry, Vol. 3, (Academic Press, New York), pp. 93-102).

Normally, regeneration will be involved in obtaining a whole plant from a transformation process. The term "regeneration" as used herein, means growing a whole plant from a plant cell, a group of plant cells, a plant part, or a plant piece (e.g., from a protoplast, calys, or a tissue part).

The foregoing methods for transformation would typically be used for producing transgenic inbred lines. Transgenic inbred lines could then be

crossed, with another (non-transformed or transformed) inbred line, in order to produce a transgenic hybrid plant. Alternatively, a genetic trait which has been engineered into a particular line using the foregoing transformation techniques could be moved into another line using traditional backcrossing techniques that are well known in the plant breeding arts. For example, a backcrossing approach could be used to move an engineered trait from a public, non-elite line into an elite line, or from a hybrid plant containing a foreign gene in its genome into a line or lines which do not contain that gene. As used herein, "crossing" can refer to a simple X by Y cross, or the process of backcrossing, depending on the context.

Parts obtained from the regenerated plant, such as flowers, pods, seeds, leaves, branches, fruit, and the like are covered by the invention, provided that these parts comprise cells which have been so transformed. Progeny and variants, and mutants of the regenerated plants are also included within the scope of this invention, provided that these parts comprise the introduced DNA sequences.

Once a transgenic plant is produced having a desired characteristic, it will be useful to propagate the plant and, in some cases, to cross to inbred lines to produce useful hybrids.

In seed propagated crops, mature transgenic plants may be self crossed to produce a homozygous inbred plant. The inbred plant produces seed containing the genes for the newly introduce trait. These seeds can be grown to produce plants that will produce the selected phenotype.

This invention further contemplates the identification of other polynucleotides encoding *Ra1* type proteins. Methods for identifying these other polynucleotides are known to those of skill in the art and will typically be based on screening for other cells which express *Ra1*. Nucleotide sequences encoding this protein are easily ascertainable to those of skill in the art through Genbank or the use of plant protein codon optimization techniques known to those of skill in the art and disclosed in the references cited herein (for example see EPO publication number 0682115A1 and Murray et al., 1989,

Nuc Acid Res., Vol. 17 No. 2, pp 447-498, "Codon Usage in Plant Genes". It is preferred to use the optimized coding sequences, for the plant recipient species. These sequences can be used not only in transgenic protocols but as tags for marker-assisted selection in plant breeding programs.

The present invention also provides antibodies capable of binding to *Ra1* from one or more selected species. Polyclonal or monoclonal antibodies directed toward part or all of a selected *Ra1* gene product may be prepared according to standard methods. Monoclonal antibodies may be prepared according to general methods of Köhler and Milstein, following standard protocols

Purified *Ra1*, or fragments thereof, may be used to produce polyclonal or monoclonal antibodies which may serve as sensitive detection reagents for the presence and accumulation of the proteins in cultured cells or tissues and in intact organisms. Recombinant techniques enable expression of fusion proteins containing part or all of a selected *Ra1*. The full length protein or fragments of the protein may be used to advantage to generate an array of monoclonal or polyclonal antibodies specific for various epitopes of the protein, thereby providing even greater sensitivity for detection of the protein.

Polyclonal or monoclonal antibodies immunologically specific for *Ra1* may be used in a variety of assays designed to detect and quantitate the proteins. Such assays include, but are not limited to, (1) immunoprecipitation followed by protein quantification; (2) immunoblot analysis (e.g., dot blot, Western blot) (3) radioimmune assays, (4) nephelometry, turbidometric or immunochromatographic (lateral flow) assays, and (5) enzyme-coupled assays, including ELISA and a variety of qualitative rapid tests (e.g., dip-stick and similar tests).

Polyclonal or monoclonal antibodies that immunospecifically interact with *Ra1* can be utilized for identifying and purifying such proteins. For example, antibodies may be utilized for affinity separation of proteins with which they immunospecifically interact. Antibodies may also be used to

immunoprecipitate proteins from a sample containing a mixture of proteins and other biological molecules.

The following examples are intended to further illustrate the invention and are not intended to limit the invention in any way. The examples and discussion herein may specifically reference maize, however the teachings herein are equally applicable to any other grass, or flowering crop.

The following examples are offered to illustrate but not limit the invention. Thus, they are presented with the understanding that various formulation modifications as well as method of delivery modifications may be made and still be within the spirit of the invention.

EXAMPLES

EXAMPLE 1

Multiple mutant alleles of *Ra1*. Seven plants expressing mutant alleles of *Ra1* were analyzed. Four novel alleles have been identified from directed tagging efforts (*Ra1-m1*, *-m2* and *-m3* from *Spm* mutagenesis, *Ra1-mum1* from *Mutator*; *Ra1-Mum1* has been renamed to *Ra1-m4*). In addition, the reference allele (*Ra1-ref*) and *Ra1-IHO* arose spontaneously, while *Ra1-RS* was provided and arose in a targeted *Spm* screen but is not due to the insertion of autonomous *Spm*. Additional putative isolates have also been reviewed, or recovered in our transposon tagging screens.

We have converged each of the verified alleles into standard inbred lines, allowing comparison of allele-specific phenotypes (Fig. 3). The relative strengths of the mutations in the allelic series are as follows:

					Ra1-ref	
weak	*Ra1-IHO*	<	*Ra1-RS*	<	*Ra1-Mum1*	**strong**
alleles					*Ra1-m1, -m2, -m3*	**alleles**

All strong alleles show branching over the length of the central spike of the tassel, nearly to the tip, and extreme branching and poor fertility in the ear. Mutant *Ra1-RS* tassels are slightly less affected; ears are highly branched but less so than for the strong alleles and correspondingly show greater female

fertility. The *Ra1-IHO* allele is unique in that it affects branching in the tassel, without affecting ear morphology (T. Berke and T. Rocheford, submitted) even when heterozygous with *Ra1-ref*. The reduction in ear and tassel phenotype across the allelic series suggests that *Ra1* mutations affect a simple locus that regulates branch programs in the tassel and ear by a single mechanism. The reported variations in *ra-ref* mutant phenotypes are thus likely the result of genetic background effects. Such background effects suggest the existence of other genes in maize whose function overlaps with that of *Ra1*, raising the possibility that *Ra1* is a member of a gene family.

The ear phenotype of *Ra1* mutants is difficult to evaluate at the mature ear stage (Postlethwait and Nelson 1964) (see Fig. 2). Initial analysis of *Ra1-ref* ears using the SEM has revealed that the silk proliferation observed in ears of *Ra1-ref* plants may not be simply due to florets present on extra branches in these genotypes. We have detected floral abnormalities, such as possible spikelet defects and extra, central carpels, that may also help explain the poor fertility associated with *Ra1-ref* (Fig. 4). Similar extra carpel initiations lead to semisterility in *zag1* and *kn1* mutants of maize (Kerstetter *et al.* 1997; Mena *et al.* 1996)

Directed transposon tagging.

The transposable element *Suppressor Mutator (Spm)* has been widely used for transposon tagging in maize. Applicants felt it was uniquely suited for tagging genes with adult phenotypes, such as *ramosa* , because it is highly mutagenic, and it generates large (early) revertant sectors and many derivative alleles. Autonomous *Spm* elements encode 2 proteins required for transposition of themselves and their deletion derivatives *dSpm (defective Spm)*.

o2-m20::Spm was used as a female parent for targeted tagging crosses (Schmidt, Burr and Burr 1987).

o2 is located 14 cM (and across the centromere) from *Ra1*. The *o2-m20::Spm* stock contained at least 2 other autonomous *Spm* elements in the cross:

o2-m20::Spm

X

o2 Ra1-ref gl1 ij1

55,000 unselected F1 kernels were sown in the field and screened at the tassel stage for highly branched tassels. Three candidates were identified and found to be heritable. Each mutation conditioned ear branching and unstable phenotypes that frequently reverted to normal in somatic sectors (Fig. 5). Each isolate showed tight linkage to *gl1+* and did not complement *Ra1-ref*. Hence, all three are allelic to *Ra1* and were named *Ra1-m1*, *Ra1-m2* and *Ra1-m3*. In the course of genetic experiments we have isolated recombinants between each *Ra1-m* allele and *o2*, as well as between *Ra1-m* loci and *gl1*. These recombinant chromosomes may ultimately be useful for verifying a clone. For all three *Ra1-m* alleles we have isolated chromosomes that are germinally revertant at *Ra1*, conditioning only normal phenotypes when homozygous or when heterozygous with *Ra1-ref*.

In similar directed tagging experiments we have utilized *Mutator* transposable elements (Robertson 1978). 25,000 plants derived from the cross:

bz1-mum9; MuDR X *o2 Ra1-ref gl1 ij1/+; bz1 sh1 wx1*

were planted in the field and screened for the *ramosa* phenotype. A single new allele, *ra-mum1*, was recovered, self-pollinated and subsequently backcrossed to the male parent. *Ra1-mum1* has been converged into B73 for 3 generations, to reduce *Mu* element copy number and randomize remaining elements in different outcross pedigrees (Martienssen *et al.* 1989). Despite extensive cosegregation analysis by conventional Southern blot (Martienssen *et al.* 1989) and AFLP-based approaches (Frey, Stettner and Gierl 1998), we have not yet

identified a potential *Mu* tag. We have element-specific probes for all of the known *Mu* elements in the lab (Chandler and Hardeman 1992). In a second screen, 24,000 F1 kernels were sown and a single putative *Ra1* plant was identified, *ra*68*. Applicants isolated DNA from 40 non-mutant siblings for immediate Southern analysis, but using *Mu1*, *Mu7*, *Mu8* and *MuDR* as probes they have not yet identified a new transposition event in *ra*68* that cosegregates with the mutation.

The genetically unstable phenotypes of the *Ra1-m* alleles strongly suggest transposon insertion. Although *Spm* was our intended mutagen, previous directed tagging strategies have unexpectedly yielded mutations tagged by transposons of a different family (McClintock, 1954; Johal and Briggs 1992; Michel *et al.* 1995; Patterson *et al.* 1995). However, the large sizes of *Ra1*-m revertant sectors are consistent with insertion of an *Spm* family member, and are inconsistent with *Mutator* insertion, which typically generates very small, late somatic sectors. We have eliminated the transposons *Bergamo* (*Bg*) (Michel *et al.* 1995) and *Ac* as a source of mutability by testcrossing our *ra* alleles with *o2-rBg* and *r1-m3::Ds* testers. These results suggest *Spm* might be involved in the *Ra1-m* alleles.

***Ra1-m2* and *Ra1-m3* genetic characterization.** To test for association of *ramosa* mutability with *Spm* activity, we first diluted *Spm* copy number by successive backcrosses with a stock that contained the *Spm* reporter *c1-m1::dSpm* and a genetically marked *Ra1-ref* chromosome.

F0 *Ra1-m +*; *c1*; +*Spm*, X *Ra1-ref gl1*; *c1-m1::dSpm*
Ra1-ref gl1 c1 *Ra1-ref gl1 c1-m1::dSpm*

non-glossy, *ra*-mutable plants glossy, *ra* plants

F1 *Ra1-m +*; *c1* ; ±*Spm* **and** *Ra1-ref gl1*; *c1* ; ±Spm
Ra1-ref gl1 c1-m1::dSpm *Ra1-ref gl1 c1-m1::dSpm*

non-glossy, *ra*, *ra*-m and N plants

glossy, *ra* plants

gl1 mutants have altered cuticular wax (glossy leaves) that can be scored easily in juvenile leaves and with careful inspection up to maturity. *c1-m1::dSpm* kernels containing *Spm* have spotted aleurones, while those lacking *Spm* have colorless aleurones. In the F1 spotted kernels and colorless kernels (if present) were sown, and non-glossy *ra*- mutable plants were selected and backcrossed once more. In some of our pedigrees, *gl* was not included, so that *ra*-mutable alleles could only be followed in mature plants (Table 1). Using this scheme, we have produced lines that segregate 1:1 for active *Spm* in *Ra1-m2* and *Ra1-m3* lines, indicating the presence of a single autonomous *Spm* element. In contrast, *Ra1-m1* lines have been recovered without *Spm*, suggesting another element may be responsible for mutability in this case.

Select backcross pedigree data are presented in Table 1, to demonstrate several points. Namely, somatic and germinal mutability is relatively high for both alleles, consistent with *Spm* family member insertion (Masson *et al.* 1987). We have never separated the *ra*-mutable phenotype from the presence of active *Spm* in our *Ra1-m2* and *Ra1-m3* backcross lines. In three independently derived *Ra1-m2* lines (lines 1, 2 and 3 in Table 1) and two *Ra1-m3* lines (lines 4 and 5 in Table 1), the single remaining *Spm* in each line is tightly linked to the *Ra1-m* locus. In several cases for *Ra1-m3*, we observed that when *Spm* transposes away a functional allele is typically left behind (lines 6 and 7), consistent with the high somatic and germinal mutability observed for *Ra1-m3*. In another case, we recovered a "change of state" derivative allele (line 3) with a weakened phenotype that retained *Spm* (McClintock 1954). We confirmed that only the tightly linked *Spm* at its original location, and not an element unlinked to *Ra1*, can confer mutability on the locus (line 8), arguing against non-autonomous *dSpm* insertion. In sum,

genetic data argue strongly for autonomous *Spm* insertion at the *Ra1* locus in the *Ra1-m2* and *Ra1-m3* alleles.

	genotype tested	*Spm* location	kernels Sown	plant phenotype	# plants	Tassel phenotype *ra*	*ra*-m	*ra*-weak	Normal	freq *ra*-m
1	*Ra1-m2* + / *Ra1-ref* *gl1*	linked	+*Spm*	non-*gl*	26	15	5	1	5	19%
				gl	2	2	0	0	0	0%
			-*Spm*	non-*gl*	2	0	0	0	2	0%
				gl	35	35	0	0	0	0%
2	*Ra1-m2* + / + +	linked	+*Spm*	can't score	14	3	8	3	0	57%
			-*Spm*	*gl*	36	0	0	0	36	0%
3	*Ra1-m2* + / + +	linked	+*Spm*	can't score	47	18	0	25	4	0%
			-*Spm*	*gl*	49	0	0	0	49	0%
4	*Ra1-m3* + / + +	linked	+*Spm*	can't score	47	28	1	0	18	2%
			-*Spm*	*gl*	43	0	0	0	43	0%
5	*Ra1-m3* + / + +	linked	+*Spm*	can't score	43	10	2	0	31	5%
			-*Spm*	*gl*	19	0	0	0	19	0%
6	*Ra1-m3* + / + +	no *Spm*	-*Spm*	can't score	43	0	0	0	43	0%
				gl						
7	*Ra1-m3* + / + +	no *Spm*	-*Spm*	can't score	47	0	0	0	47	0%
				gl						
8	*Ra1-m3* + / + +	unlinked	+*Spm*	can't score	41	24	0	0	17	0%
			-*Spm*	*gl*	40	20	0	0	20	0%

Table 1. The data show progeny results when the indicated genotypes are testcrossed by a *Ra1-ref gl1; c1-m1::dSpm* stock.

Southern analysis of the *Ra1-m2* and *Ra1-m3* single-*Spm* lines.

SalI restriction enzyme sites in the *Spm* promoter region are methylated when the element is inactive, and unmethylated when it is active (Banks, Masson and Fedoroff 1988). Thus, genomic DNA blots using *SalI* were used to verify that a single active *Spm* segregates in each line (not shown). Moreover, genomic DNA blots using methylation insensitive restriction enzymes show that *Ra1-m2* and *Ra1-m3* contain different *Spm* elements, ruling out the possibility of a closely linked element pre-existing prior to mutagenesis. For *Ra1-m2*, the co-segregating fragment resolves as a 16 kb *HinDIII* fragment or a 5 kb *EcoRI* fragment; for *Ra1-m3* we have identified a 3.6 kb *HinDIII* fragment (Fig. 6). These genetic and molecular analyses show that *Ra1-m2* and *Ra1-m3* are each likely caused by insertion of a unique, autonomous *Spm* element.

<u>Molecular genetics of Ramosa 1.</u>

The *Ra1* locus. Isolate molecular clones have been isolated of the *Spm* transposons linked to the *Ra1-m2* and *Ra1-m3* alleles, for which we have the best cosegregation data. PCR approaches have proved problematic due to the high rates of *Spm* transposition. However, we have extensive experience in the construction and screening of phage libraries (Hake, Vollbrecht and Freeling 1989; Han, Coe and Martienssen 1992; Martienssen *et al.* 1989), and we will clone the 4.9 kb *EcoRI* fragment from *Ra1-m2* using a standard lambda replacement vector (lambda-ZAP II, Stratagene) and *Spm* probes. We will use size-fractionated, gel purified genomic DNA. This clone will contain approximately 3.5 kb of genomic DNA flanking the 5' end of *Spm*, which we will use as a hybridization probe on the same DNA gel blots as before. This will require first identifying non-repetitive portions of the clone, which we will do through a combination of sequencing and reverse Southern hybridization in which total genomic DNA is labeled and hybridized to restriction digests of the cloned DNA (Martienssen *et al.* 1989). If the entire clone is repetitive, we will

isolate the 16kb *HinDIII* fragment from *Ra1-m2* instead. If *Ra1-m2* doesn't yield a *Ra1* clone then our second choice will be the 3.5 kb *HinDIII* fragment from *Ra1-m3*, which we will clone by size-selected gel purification and plasmid cloning (Colasanti, Yuan and Sundaresan 1998).

Several means are available to us to prove that an isolated clone actually corresponds to the *Ra1* locus. One powerful test involves comparing the genomic DNA of multiple mutant alleles with that of their progenitors and with derivative, intragenic revertants. We have progenitor strains for each of the five alleles we isolated by transposon tagging, and we have multiple revertant derivatives from the three *Spm*-induced alleles. We have already isolated DNA from many such individuals for co-segregation analysis, and so DNA gel blot analysis will proceed rapidly. We will also use RNA expression analysis (RNA gel blots or RT-PCR, as necessary) on similar mutant-progenitor-derivative series, once a transcription unit has been defined as described below. Once we have a clone, we will define the locus. DNA gel blot analyses will reveal any gross alterations in different alleles, which will help determine whether or not *Ra1* is a complex locus, although as described above we expect a simple structure based on phenotypes in the allelic series. Characterization of the *Ra1* locus will then focus on the inbred line B73, for which several libraries are available and into which we have introgressed several of our mutant alleles. Genomic clones spanning the region of interest will be isolated from a *Sau3A* partial-digest B73 genomic DNA library (gift of S. Hake and B. Veit). A few kb around the *Spm* insertion sites will be sequenced, using the structure of different alleles as a guide to defining the most interesting regions.

We expect the *Ra1* gene product to be expressed in both the tassel and the ear, at about the time when spikelet pair meristems are initiated. We will attempt to identify candidate, transcribed regions from the sequence, using BLAST searches as well as a suite of gene modeling programs that we have previously trained on maize and other cereal genomes ((Rabinowicz *et al.* 1999); L. Stein, unpublished). In particular, we will search our sequence for

matches to expressed sequence tags (ESTs) from immature inflorescences (Walbot et al., unpublished). We will probe RNA gel blots with fragments from the locus, using poly A+ RNA from mutant and normal immature inflorescences. If RNA gel blots do not reveal any transcribed regions, we will use RT-PCR, as has proven necessary for detecting rare RNAs in the tassel (DeLong, Calderon-Urrea and Dellaporta 1993). Once the transcript(s) are identified, we will isolate and sequence cDNA clones using the same probes and amplified cDNA libraries constructed from immature ears of inbred B73 (Kerstetter *et al.* 1994). For most genes, cDNA library screening is sufficiently sensitive to obtain clones. If the transcript is expressed at very low levels, we will use primers located within the transcribed region for 5' and 3' RACE PCR (Settles *et al.* 1997). All DNA sequencing will be done at the Cold Spring Harbor genome center in collaboration with Dick McCombie (e.g. (Springer *et al.* 1995)). Molecular characterization will include wild-type and mutant alleles, with emphasis on at least two of the strong alleles as well as two weak alleles (*Ra1-RS* and *Ra1-TR*) which may prove especially informative for understanding the mechanism of gene product function.

EXAMPLE 2

The Ramosa 1 gene modifies plant architecture.

Plant architecture is defined by the relative placement of leaves, flowers and branches on the stem. Highly branched inflorescences, such as in tomato, have many more flowers per plant and yield a high quantity of fruit. However, they also provide considerable shade, reducing the density at which plants can be grown. Modern varieties of hybrid maize for example are grown at the very high densities, but have few or no branches either in the flower or in the stem. The branching pattern of crop plant architecture is thus one of the primary influences on yield, and the ability to manipulate this key trait is likely to significantly impact global plant productivity.

We have isolated a gene from maize called Ramosa 1. This gene corresponds to a classical mutation in maize that was first identified in the

early 1900s and thought to correspond to a new species. Subsequent work showed that the mutant variety was due to a single mutation that mapped to chromosome 7, the Ramosa 1 gene.

Ramosa 1 mutant plants have highly branched tassels and ears. Posthlewaite and Nelson (1958) interpreted this phenotype as a conversation of second order into first order branches, placing it at the core of the branching pathway that determines plant architecture in maize. Mutant tassels are branched from base to tip, while wild-type tassels only have branches at the base. Mutant ears are also highly branched, resembling those of sorghum and millet rather than those of maize which are of course unbranched.

4 new alleles of the Ramosa 1 gene were identified by crossing homozygous mutants to stocks carrying the *Suppressor-mutator* transposon, and screening roughly 70,000 progeny plants for those with branched tassels (McClintock, 1953). These alleles were highly mutable: both male and female inflorescences were partly branched mosaics, and both mutant and normal progeny were recovered when homozygous loss of function mutants were self pollinated. The mutants were backcrossed to a ramosa/Spm tester to reduce the number of transposons per plant, and co-segregating transposons were identified in each case. These allowed the gene to be isolated molecularly. Ramosa 1 was found to encode a zinc finger protein approximately 42% identical to the SUPERMAN gene of *Arabidopsis thaliana*. The *Arabidopsis* gene has a distinct phenotype, effecting floral determination but not branching.

The ramosa gene is a modifier of plant architecture. For example, inhibition of expression via antisense or RNAi would be expected to increase branching. This would be expected to increase the yield of fruit and seed per plant. In addition, highly branched tassels would be expected to have increased pollen shed resulting in greater fertility for use in hybrid corn production as well as increased yield.

Increasing Ramosa 1 expression, on the other hand, would be expected to reduce branching. This would be a crucial step in transforming primitive

crops such as millet and sorghum into higher yielding derivatives with unbranched maize-like ears.

Cloning the *Ra1* gene

According to the invention the *Spm*-hybridizing, 5 kb *Eco*RI fragment in *Ra1-m2* was cloned by constructing a size-selected lambda phage library, as described in Example 1. Based on reverse Southern hybridization, a 2.4 kb sequence adjacent to the *Spm* insertion was single copy, and was selected as a probe ("the *Kpn* fragment") back to DNA gel blots containing genomic DNA from other mutant *Ra1* alleles and their progenitors. The single *Spm* in the *Ra1-m3* chromosome was inserted in nearly the same place as the *Spm* in *Ra1-m2*, but in opposite orientation. Germinal revertants of *Ra1-m2* and *Ra1-m3* lack their respective insertion. The chromosome containing the *Ra1-Mum1* allele also contained an *Spm* insertion that was not present in its progenitor, placing the three insertions within a 700 nucleotide region. The *Kpn* fragment hybridizes to an RNA of 850 nucleotides that is present in tassels but not vegetative tissues. This message is expressed at normal levels in *Ra1-Ref* tassels; although *Ra1-Ref* contains no obvious rearrangements detectable at the level of DNA gel blots, sequence analysis suggests it contain a point mutation. We determined the 5' and 3' ends of this transcription unit by RACE experiments, which delineated a 692 or 741 nucleotide mRNA (two polyadenylation sites were encountered with equal frequency) that encodes a 175 amino acid protein. This gene contains no introns, and the original genomic clone contains the transcribed region plus 341 bp upstream and 2.6 kb downstream. See Figures 7, 9, 11, and 12.

The *Spm* insertions in *Ra1-m2* and *Ra1-m3* are upstream of the transcription start, while the *Spm* in *Ra1-Mum1* interrupts the coding sequence. These locations correlate well with the observed high mutability of *Ra1-m2* and *Ra1-m3*, and low mutability of *Ra1-Mum1*. This combination of coincident insertions in three independent alleles and a point mutation in a fourth all mapping to the same transcription unit, and germinal reversion

associated with transposon excision, shows that applicants have identified and cloned the *Ra1* gene.

_Ra1_ encodes a zinc finger protein similar to the SUPERMAN gene of Arabidopsis

Only the transcribed region of the genomic clone returned significant similarity in BLAST searches, to a family of putative TFIIIa-type zinc finger proteins that is specific to plants. The best hit was the SUPERMAN (SUP) gene of *Arabidopsis*, followed by a group of five zinc-finger genes, all from *Arabidopsis*. Sequences producing significant alignments are shown in figure 8.

The next best 36 alignments have relatively high E values (> e-4) yet involve the zinc finger region exclusively, indicating *Ra1* contains little homology to other proteins currently in the databases. Similarity between full-length *RA1* and SUP proteins includes their single zinc finger and an adjacent, short proline-rich motif, as well as a dozen or so residues at the carboxy terminus. SUP encodes a polyserine stretch just C-terminal to the zinc finger, while an analogous region is encoded just N-terminal in *Ra1*.

These protein sequences fall into a family of plant zinc finger proteins known as the EPF type, defined by the plant-specific, highly conserved predicted alpha-helical motif QALGGH within the finger region (ref Takatsuji review). Interestingly, the *RA1* protein contains a novel Ala→Gly within this motif, resulting in a sequence that is not present in the databases. The encoding of Gly in this position in all ten wild type *Ra1* alleles we have sequenced thus far, raises the possibility that *Ra1* produces a reduced function gene product. Reduced function could explain the few-branched morphology of the maize tassel, intermediate between unbranched (as in wheat) and fully branched (as in sorghum or *Ra1* mutants). The *Ra1-Ref* mutant allele encodes two further amino acid differences (QGLEGN) within this region. The nonconservative Gly→Glu amino acid difference, plus or minus the conservative His→Asn difference, is the likely cause of the *Ra1-Ref* lesion. The *Spm* in *Ra1-Mum1* is

inserted adjacent to but outside of the proline-rich segment. Since *Ra1-Mum1* shows only rare somatic mutability in plants that contain *Spm* and yet exhibits excision from the locus on genomic DNA blots, this insertion likely defines a region of the protein that is sensitive to point mutations.

There are at least 40 EPF type zinc finger proteins identified in the *Arabidopsis* genome so far. Hence, we expect *Ra1* to be part of a gene family, as we speculate in our original proposal. Only four EPF sequences from maize are in the public databases, indicating that this gene family has been relatively unexplored by functional analyses to date.

The following references are hereby incorporated herein by reference.

Allsopp, A., 1967 Heteroblastic development in vascular plants. Adv Morph **6:** 127-171.

Athma, P., E. Grotewold and T. Peterson, 1992 Insertional mutagenesis of the maize P gene by intragenic transposition of Ac. Genetics **131:** 199-209.

Banks, J. A., P. Masson and N. Fedoroff, 1988 Molecular mechanisms in the developmental regulation of the maize Suppressor-mutator transposable element. Genes And Development **2:** 1364-80.

Barton, M., and R. Poethig, 1993 Formation of the shoot apical meristem in *Arabidopsis*: an analysis of development in the wild type and the *shoot meristemless* mutant. Development **119:** 823-31.

Bradley, D., R. Carpenter, L. Copsey, C. Vincent, S. Rothstein and E. Coen, 1996 Control of inflorescence architecture in *Antirrhinum*. Nature **379:** 791-797.

Bradley, D., O. Ratcliffe, C. Vincent, R. Carpenter and E. Coen, 1997 Inflorescence commitment and architecture in *Arabidopsis*. Science **275:** 80-83.

Burr, B., F. A. Burr, K. H. Thompson, M. C. Albertson and C. W. Stuber, 1988 Gene mapping with recombinant inbreds in maize. Genetics **118:** 519-526.

Cacharr n, J., H. Saedler and G. Theissen, 1999 Expression of MADS box genes ZMM8 and ZMM14 during inflorescence development of Zea mays

discriminates between the upper and the lower floret of each spikelet. Dev Genes Evol **209:** 411-20.

Chandler, V., and K. Hardeman, 1992 The Mu elements of Zea mays. Adv. Genet. **30:** 77-122.

Cheng, P. C., R. I. Greyson and D. B. Walden, 1983 Organ initiation and the development of unisexual flowers in the tassel and ear of *Zea mays.* Amer J Bot **70:** 450-462.

Chomet, P., 1994 Transposon tagging with *Mutator.* In *The Maize Handbook,* edited by M. Freeling, and V. Walbot, pp. 243-249. Springer-Verlag, New York.

Chuck, G., R. Meeley and S. Hake, 1998 The control of maize spikelet meristem identity by the *APETALA-2*-like gene *indeterminate spikelet1.* Genes & Dev **12:** 1145-1154.

Clark, S., S. Jacobsen, J. Levin and E. Meyerowitz, 1996 The *CLAVATA* and *SHOOTMERISTEMLESS* loci competitively regulate meristem identity in *Arabidopsis.* Development **122:** 1567-1575.

Clark, S. E., M. P. Running and E. M. Meyerowitz, 1993 *CLAVATA1,* a regulator of meristem and flower development in *Arabidopsis.* Development **119:** 397-418.

Clark, S. E., M. P. Running and E. M. Meyerowitz, 1995 *CLAVATA3* is a specific regulator of shoot and floral meristem development affecting the same processes as *CLAVATA1.* Development **121:** 2057-2067.

Clifford, H., 1987 Spikelet and floral morphology. In *Grass systematics and evolution,* edited by T. Soderstrom, K. Hilu, C. Campbell and M. Barkworth, pp. 21-30. Smithsonian Institution Press, Washington, D.C.

Coen, E. S., J. M. Romero, S. Doyle, R. Elliot, G. Murphy and R. Carpenter, 1990 *FLORICAULA*: a homeotic gene required for flower development in *Antirrhinum majus.* Cell **63:** 1311-1322.

Colasanti, J., Z. Yuan and V. Sundaresan, 1998 The *indeterminate* gene encodes a zinc finger protein and regulates a leaf-generated signal required for the transition to flowering in maize. Cell **93:** 593-603.

Collins, G., 1917 Hybrids of *Zea ramosa* and *Zea tunicata*. J Agric Res **9:** 383-413.

Cubas, P., N. Lauter, J. Doebley and E. Coen, 1999 The TCP domain: a motif found in proteins regulating plant growth and development. Plant Journal **18:** 215-22.

Das, L., and R. Martienssen, 1995 Site-selected transposon mutagenesis at the *hcf106* locus in maize. Plant Cell **7:** 287-294.

Dawe, R. K., and M. Freeling, 1990 Clonal analysis of the cell lineage in the male flower of maize. Dev. Biol. **142:** 233-245.

DeLong, A., A. Calderon-Urrea and S. L. Dellaporta, 1993 Sex determination gene TASSELSEED2 of maize encodes a short-chain alcohol dehydrogenase required for stage-specific floral organ abortion. Cell **74:** 757-68.

Doebley, J., A. Stec and L. Hubbard, 1997 The evolution of apical dominance in maize. Nature **386:** 485-8.

Evans, M., and R. Poethig, 1995 Gibberellins promote vegetative phase change and reproductive maturity in maize. Plant Physiology **108:** 475-487.

Ferrandiz, C., Q. Gu, R. Martienssen and M. Yanofsky, 2000 Redundant regulation of meristem identity and plant architecture by *FRUITFULL*, *APETALA1* AND *CAULIFLOWER*. Development : in press.

Frey, M., M. Stettner and A. Gierl, 1998 A general method for gene isolation in tagging approaches: amplification of insertion mutagenized sites (AIMS). Plant J **13:** 717-721.

Gernert, W., 1912 A new subspecies of *Zea mays* L. Amer Naturalist **46:** 616-622.

Gu, Q., C. Ferrandiz, M. F. Yanofsky and R. Martienssen, 1998 The *FRUITFULL* MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. Development **125:** 1509-17.

Hake, S., and M. Freeling, 1986 Analysis of genetic mosaics shows that the extra epidermal cell divisions in *Knotted* mutant maize plants are induced by adjacent mesophyll cells. Nature (London) **320:** 621-623.

Hake, S., E. Vollbrecht and M. Freeling, 1989 Cloning *Knotted*, the dominant morphological mutant in maize using *Ds2* as a transposon tag. EMBO Journal **8:** 15-22.

Han, C.-d., E. J. Coe and R. Martienssen, 1992 Molecular cloning and characterization of *iojap* (*ij*), a pattern striping gene of maize. EMBO J **11:** 4037-4046.

Hansen, J. D., J. Pyee, Y. Xia, T. J. Wen, D. S. Robertson, P. E. Kolattukudy, B. J. Nikolau and P. S. Schnable, 1997 The glossy1 locus of maize and an epidermis-specific cDNA from Kleinia odora define a class of receptor-like proteins required for the normal accumulation of cuticular waxes. Plant Physiology **113:** 1091-100.

Huala, E., and I. M. Sussex, 1992 *LEAFY* interacts with floral homeotic genes to regulate *Arabidopsis* floral development. Plant Cell **4:** 901-913.

Irish, E. E., 1998 Additional vegetative growth in maize reflects expansion of fates in preexisting tissue, not additional divisions by apical initials. Developmental Biology **197:** 198-204.

Jackson, D., 1991 In situ hybridization in plants. In *Molecular Plant Pathology: A Practical Approach*, edited by D. J. Bowles, S. J. Gurr and M. McPherson, pp. 163-174. Oxford Univ. Press, Oxford.

Jackson, D., B. Veit and S. Hake, 1994 Expression of maize *KNOTTED1* related homeobox genes in the shoot apical meristem predicts patterns of morphogenesis in the vegetative shoot. Development **120:** 405-413.

Jofuku, K. D., B. G. den Boer, M. Van Montagu and J. K. Okamuro, 1994 Control of Arabidopsis flower and seed development by the homeotic gene *APETALA2*. Plant Cell **6:** 1211-25.

Johal, G., and S. Briggs, 1992 Reductase activity encoded by the *HM1* disease resistance geen in maize. Science **258:** 985-987.

Johri, M., and E. Coe, Jr, 1982 Clonal analysis of corn plant develpoment. I. The development of the tassel and the ear shoot. Dev Biol **97:** 154-172.

Kellogg, E., and J. Birchler, 1993 Linking phylogeny and genetics: *Zea mays* as a tool for phylogenetic studies. Syst Biol **42:** 425-439.

Kellogg, E., and H. Shaffer, 1993 Model organisms in evolutionary studies. Syst Biol **42:** 409-414.

Kempton, J., 1921 Inheritance of ramose inflorescence in maize. USDA Bull **971:** 1-20.

Kerstetter, R., E. Vollbrecht, B. Lowe, B. Veit, J. Yamaguchi and S. Hake, 1994 Sequence analysis and expression patterns divide the maize *knotted1*-like homeobox genes into two classes. Plant Cell **6:** 1877-1887.

Kerstetter, R. A., D. Laudencia-Chingcuanco, L. G. Smith and S. Hake, 1997 Loss of function mutations in the maize homeobox gene, *knotted1*, are defective in shoot meristem maintenance. Development **124:** 3045-3054.

Langdale, J., E. Irish and T. Nelson, 1994 Action of the *Tunicate* locus on maize floral development. Dev Genet **15:** 176-187.

Laux, T., K. F. X. Mayer, J. Berger and G. Jurgens, 1996 The *WUSCHEL* gene is required for shoot and floral meristem integrity in *Arabidopsis*. Development **122:** 87-96.

Lejeune, P., and G. Bernier, 1996 Effect of environment on the early steps of ear initiation in maize (*Zea mays L.*). Plant, Cell & Env **19:** 217-224.

Leyser, H. M. O., and I. J. Furner, 1992 Characterisation of three shoot apical meristem mutants of Arabidopsis thaliana. Development **116:** 397-403.

Liljegren, S. J., C. Gustafson-Brown, A. Pinyopich, G. S. Ditta and M. F. Yanofsky, 1999 Interactions among APETALA1, LEAFY, and TERMINAL FLOWER1 specify meristem fate. Plant Cell **11:** 1007-18.

Lucas, W. J., S. Bouche-Pillon, D. P. Jackson, L. Nguyen, L. Baker, B. Ding and S. Hake, 1995 Selective trafficking of KNOTTED1 homeodomain protein and its mRNA through plasmodesmata. Science **270:** 1980-1983.

Maes, T., M. Van Montagu and T. Gerats, 1999 The inflorescence architecture of *petunia hybrida* is modified by the *Arabidopsis thaliana AP2* gene. Dev Genet **25:** 199-208.

Martienssen, R., 1997 The origin of maize branches out. Nature **386:** 443, 445.

Martienssen, R., and L. Dolan, 1998 Patterns in vegetative development. In *Arabidopsis*, edited by M. Anderson, and J. Roberts, pp. 262-289. Sheffield University Press, Sheffield.

Martienssen, R., and V. Irish, 1999 Copying our ABCs: the role of gene redundancy in interpreting genetic hierarchies. Trends Genet **15:** 435-437.

Martienssen, R. A., A. Barkan, M. Freeling and W. Taylor, 1989 Molecular cloning of a maize gene involved in photosynthetic membrane organization that is regulated by Robertson's Mutator. EMBO J. **8:** 1633-1639.

Masson, P., R. Surosky, J. A. Kingsbury and N. V. Fedoroff, 1987 Genetic and molecular analysis of the Spm-dependent a-m2 alleles of the maize a locus. Genetics **117:** 117-37.

McClintock, B., 1954 Mutations in maize and chromosomal aberrations in *Neurospora*. Carnegie Institute of Washingotn Year Book **53.**

McConnell, J., and M. Barton, 1995 Effect of mutations in the *PINHEAD* gene of *Arabidopsis* on the formation of shoot apical meristems. Dev. Genet. **16:** 358-366.

Mena, M., B. A. Ambrose, R. B. Meeley, S. P. Briggs, M. F. Yanofsky and R. J. Schmidt, 1996 Diversification of C-function activity in maize flower development. Science **274:** 1537-40.

Michel, D., H. Hartings, S. Lanzini, M. Michel, M. Motto, G. Riboldi, F. Salamini and H.-P. Doring, 1995 Insertion mutations at the maize *opaque2* locus induced by transposable element families *Ac, En/Spm* and *Bg*. Mol Gen Genet **248:** 287-292.

Moreno, M. A., J. Chen, I. Greenblatt and S. L. Dellaporta, 1992 Reconstitutional mutagenesis of the maize P gene by short-range Ac transpositions. Genetics **131:** 939-56.

Moussian, B., H. Schoof, A. Haecker, G. Jurgens and T. Laux, 1998 Role of the *Zwille* gene in the regulation of central shoot meristem cell fate during *Arabidopsis* embryogenesis. EMBO J **17:** 1799-1809.

Neuffer, M., E. Coe and S. Wessler, 1997 *Mutants of Maize*. Cold Spring Harbor Laboratory Press, Plainview.

Okamuro, J. K., W. Szeto, C. Lotys-Prass and K. D. Jofuku, 1997 Photo and hormonal control of meristem identity in the Arabidopsis flower mutants apetala2 and apetala1. Plant Cell **9:** 37-47.

Patterson, G., K. Kubo, T. Shroyer and V. Chandler, 1995 Sequences required for paramutation of the maize *b* gene map to a region containing the promoter and upstream sequences. Genetics **140:** 1389-1406.

Phinney, B. O., and C. A. West, 1960 Gibberellins and the growth of flowering plants. In *Developing Cell Systems and Their Control*, edited by D. Rudnick, pp. 71-92. Ronald Press, New York.

Pneuli, L., L. Carmel-Goren, D. Hareven, T. Gutfinger, J. Alvarez, M. Ganal, D. Zamir and E. Lifschitz, 1998 The *SELF-PRUNING* gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is the ortholog of *CEN* and *TFL1*. Development **125:** 1979-89.

Poethig, R., 1990 Phase change and the regulation of shoot morphogenesis in plants. Science **250:** 923-930.

Postlethwait, S. N., and O. E. Nelson, 1964 Characterization of development in maize through the use of mutants. I. The *polytypic* (*Pt*) and *ramosa-1* (*Ra1*) mutants. Amer. J. Bot. **51:** 238-243.

Putterill, J., F. Robson, K. Lee, R. Simon and G. Coupland, 1995 The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarity to zinc finger transcriptin factors. Cell **80:** 847-857.

Rabinowicz, P., K. Schutz, N. Dedhia, C. Yordan, L. Parnell, L. Stein, W. McCombie and R. Martienssen, 1999 Differential methylation of genes and retrotransposons allows shotgun sequencing of the maize genome. Nat Genet **23:** 305-308.

Ratcliffe, O., I. Amaya, C. Vincent, S. Rothstein, R. Carpenter, E. Coen and D. Bradley, 1998 A common mechanism controls the life cycle and architecture of plants. Development **125:** 1609-1615.

Robertson, D. S., 1978 Characterization of a mutator system in maize. Mut. Res. 51: 21-28.

Schmidt, R., F. Burr and B. Burr, 1987 Transposon tagging and molecular analysis of the maize regulatory locus *opaque-2*. Science 238: 96-963.

Schmitz, G., and K. Theres, 1999 Genetic control of branching in Arabidopsis and tomato. Curr Opin Plant Biol 2: 51-5.

Settles, A. M., A. Yonetani, A. Baron, D. R. Bush, K. Cline and R. Martienssen, 1997 Sec-independent protein translocation by the maize Hcf106 protein. Science 278: 1467-70.

Shannon, S., and D. R. Meeks-Wagner, 1991 A mutation in the Arabidopsis *TFL1* gene affects inflorescence meristem development. Plant Cell 3: 877-892.

Sinha, N., and S. Hake, 1990 Mutant characters of knotted maize leaves are determined in the innermost tissue layers. Developmental Biology 141: 203-10.

Springer, P. S., W. R. McCombie, V. Sundaresan and R. A. Martienssen, 1995 Gene trap tagging of *PROLIFERA*, an essential MCM2-3-5-like gene in *Arabidopsis*. Science 268: 877-80.

Steeves, T. A., and I. M. Sussex, 1989 Embryogenesis: beginnings of development. In *Patterns in Plant Development*, pp. 6-25. Cambridge University Press, Cambridge.

Sundaresan, V., P. Springer, T. Volpe, S. Haward, J. D. G. Jones, C. Dean, H. Ma and R. Martienssen, 1995 Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. Genes & Development 9: 1797-1810.

Sussex, I. M., 1989 Developmental programming of the shoot meristem. Cell 56: 225-229.

Telfer, A., and R. S. Poethig, 1998 *HASTY*: a gene that regulates the timing of shoot maturation in *Arabidopsis thaliana*. Development 125: 1889-1898.

Veit, B., R. J. Schmidt, S. Hake and M. F. Yanofsky, 1993 Maize floral development-- new genes and old mutants. Plant Cell 5: 1205-1215.

Vollbrecht, E., B. Veit, N. Sinha and S. Hake, 1991 The developmental gene *Knotted-1* is a member of a maize homeobox gene family. Nature **350:** 241-243.

Walsh, J., and M. Freeling, 1999 The *liguleless2* gene of maize functions during the transition from the vegetative to the reproductive shoot. Plant J **19:** 489-495.

Weigel, D., J. Alvarez, D. R. Smyth, M. F. Yanofsky and E. M. Meyerowitz, 1992 LEAFY controls floral meristem identity. Cell **69:** 843-859.

Weil, C. F., S. Marillonnet, B. Burr and S. R. Wessler, 1992 Changes in state of the Wx-m5 allele of maize are due to intragenic transposition of Ds. Genetics **130:** 175-85.

As can be seen from the foregoing the invention accomplishes at least all of its objectives.

What is claimed is:

1. A purified and isolated nucleotide sequence which encodes upon expression an *Ra1* protein.

2. The sequence of claim 1 wherein said sequence is isolated from maize.

3. The sequence of claim 1 wherein said sequence is isolated from sugar cane.

4. The sequence of claim 1 wherein said sequence is isolated from teosinte.

5. The nucleotide sequence of claim 1 wherein said sequence comprises a sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, and SEQ ID NO:5 and acts to control branching in a plant.

6. An expression construct comprising: a nucleotide sequence according to claim 1, operatively linked to a regulatory region capable of directing expression of a protein in a plant cell.

7. A vector capable or transforming or transfecting a host cell, said vector comprising an expression construct according to claim 6.

8. The vector of claim 7 wherein said vector is a plasmid based vector.

9. The vector of claim 7 wherein said vector is a viral based vector.

10. A prokaryotic or eucaryotic host cell transformed or transfected with a vector according to claim 7.

11. The host cell of claim 10 wherein said cell is a plant cell.

12. A *Ra1* protein which exhibits the following characteristics: a zinc finger protein transcription factor capable of influencing meristem identity and branch development in plants.

13. The protein of claim 12 wherein said protein is from maize.

14. The protein of claim 12 wherein said protein is from sugar cane.

15. The protein of claim 12 wherein said protein is from teosinte.

16. The protein of claim 12 wherein said protein comprises an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4 and SEQ ID NO:6.

17. The protein of claim 12 wherein said protein is expressed in a plant cell.

18. A method for decreasing branching in a plant comprising: introducing to a plant cell a genetic construct comprising a nucleotide sequence which encodes an *Ra1* protein, said nucleotide sequence operably linked to promoter and regulatory regions capable of inducing expression in a plant cell.

19. The method of claim 18 wherein said meristem identity is altered.

20. The method of claim 19 wherein said altered meristem identity causes meristem conversion of indeterminate to determinate plant type.

21. The method of claim 18 wherein said plant is maize.

22. A method of increasing branching and flowering to improve yield in plants comprising: inhibiting the expression of nucleotide sequence which encodes upon expression an *Ra1* protein.

23. The method of claim 22 wherein said inhibition is by antisense.

24. The method of claim 23 wherein inhibition is by co-suppression.

25. The method of claim 22 wherein said inhibition is by homologous recombination.

26. A *Ra1* mutant which increases branching in plants comprising *Ra1*-m2.

27. A *Ra1* mutant which increases branching in plants comprising Ra1-m3.

28. A *Ra1* mutant which increases branching in plants comprising *Ra1*-Mum1.

29. A *Ra1* mutant which increases branching in plants comprising *Ra1*-ref.

30. A zinc finger transcription factor capable of influencing branching in plants said factor comprising: an amino acid sequence of SEQ ID NO:2 including its conservatively modified variants, which conserve function.

31. A method of increasing inflorescence number in plants comprising: introducing to a plant cell a genetic construct, said construct comprising a nucleotide sequence which encodes upon expression a *Ra1* protein, said sequence operably linked to promoter and regulatory regions capable of causing expression in a plant cell.

32. A method of identifying genes in plant species which regulate branching and plant architecture comprising: screening the genome of said plant species for a sequence that is homologous to SEQ ID NO:1 or a region of at least 100 bases thereof.

33. A gene sequence identified by the method of claim 32.

34. A protein encoded by the sequence of claim 33.

35. An antibody which is immunologically specific for one or more epitopes of *Ral* protein.

36. The antibody of claim 35 wherein said antibody is polyclonal.

37. The antibody of claim 35 wherein said antibody is monoclonal.

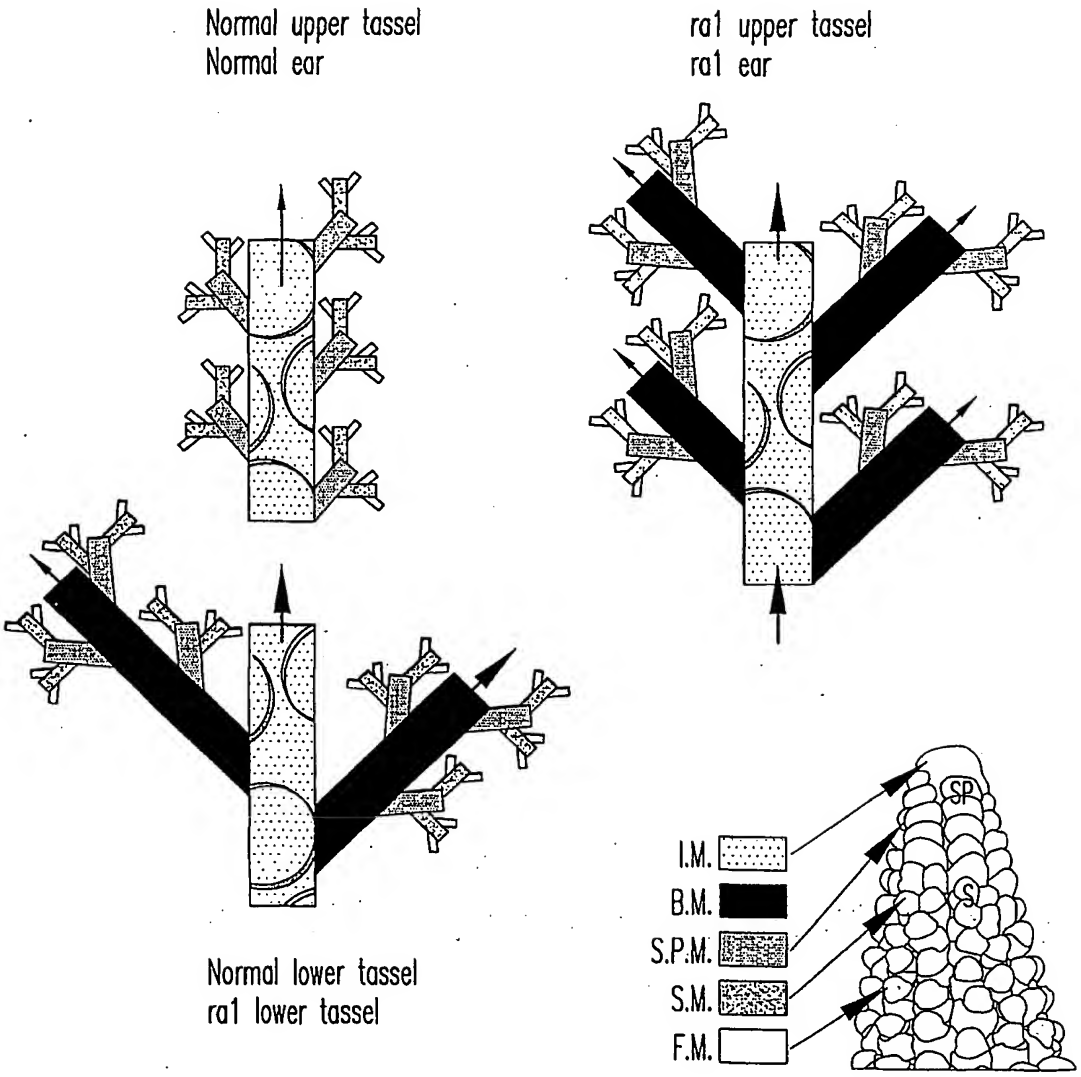

Normal upper tassel
Normal ear
ra1 upper tassel
ra1 ear
Normal lower tassel
ra1 lower tassel
I.M.
B.M.
S.P.M.
S.M.
F.M.
SP
S

Fig. 1

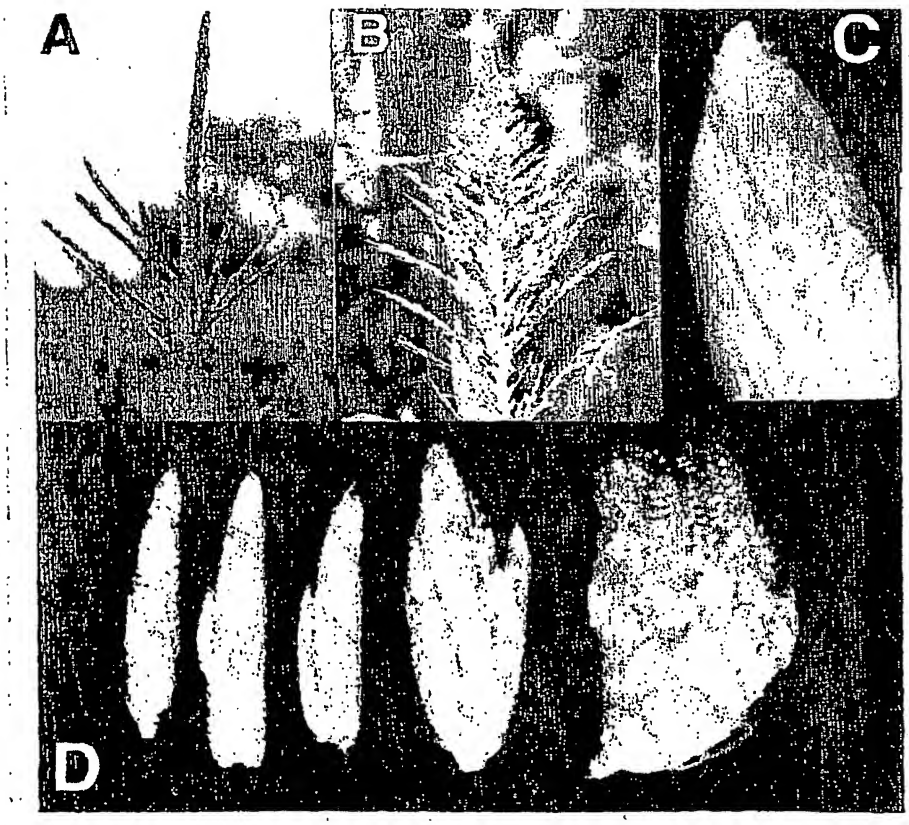


Fig. 2

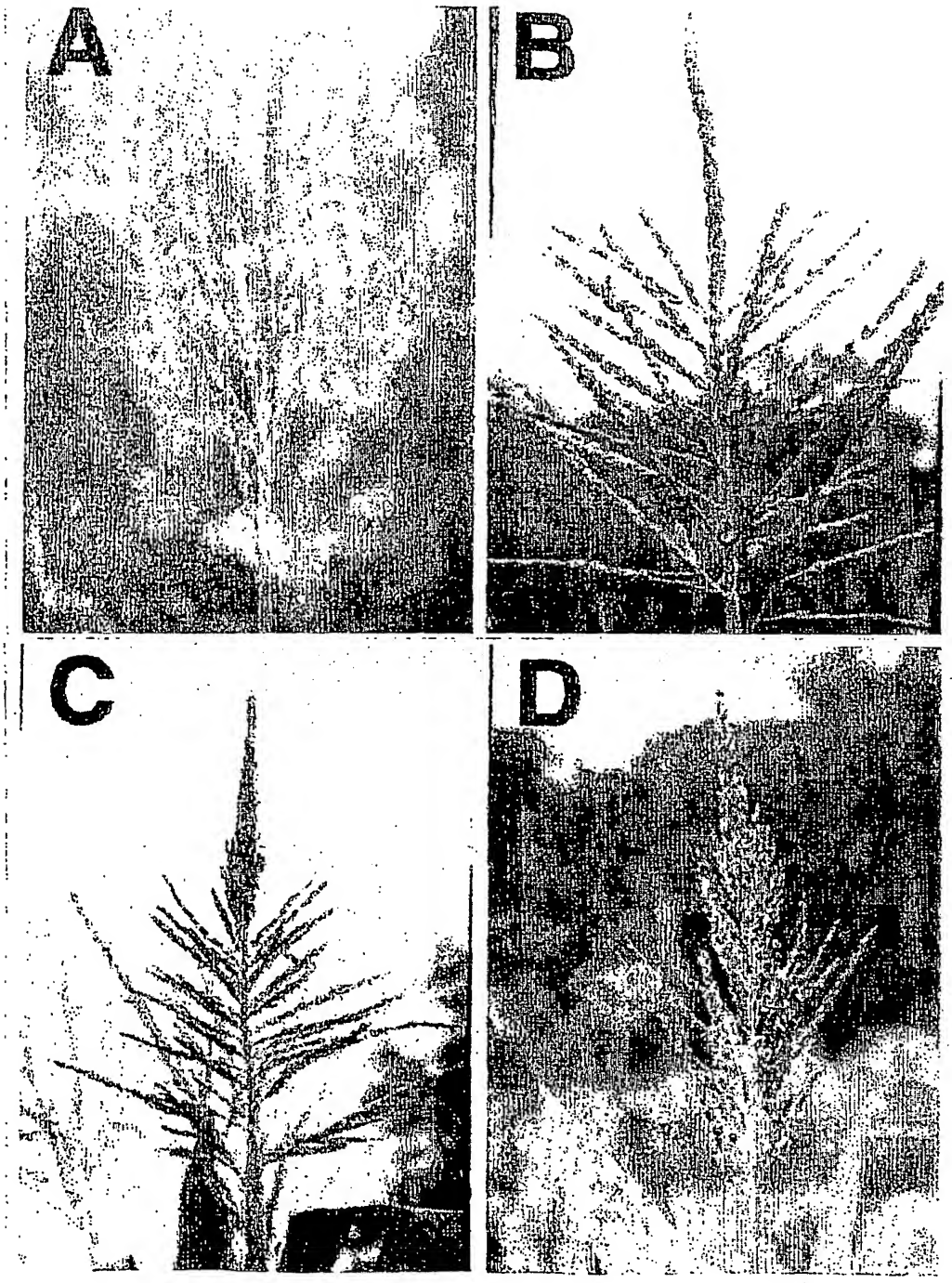


Fig. 3

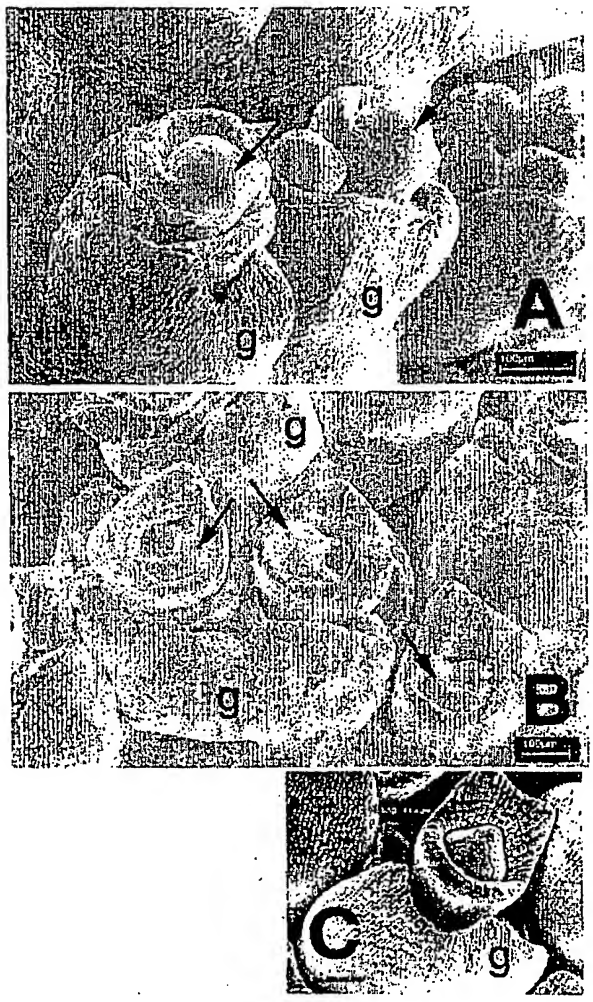


Fig. 4

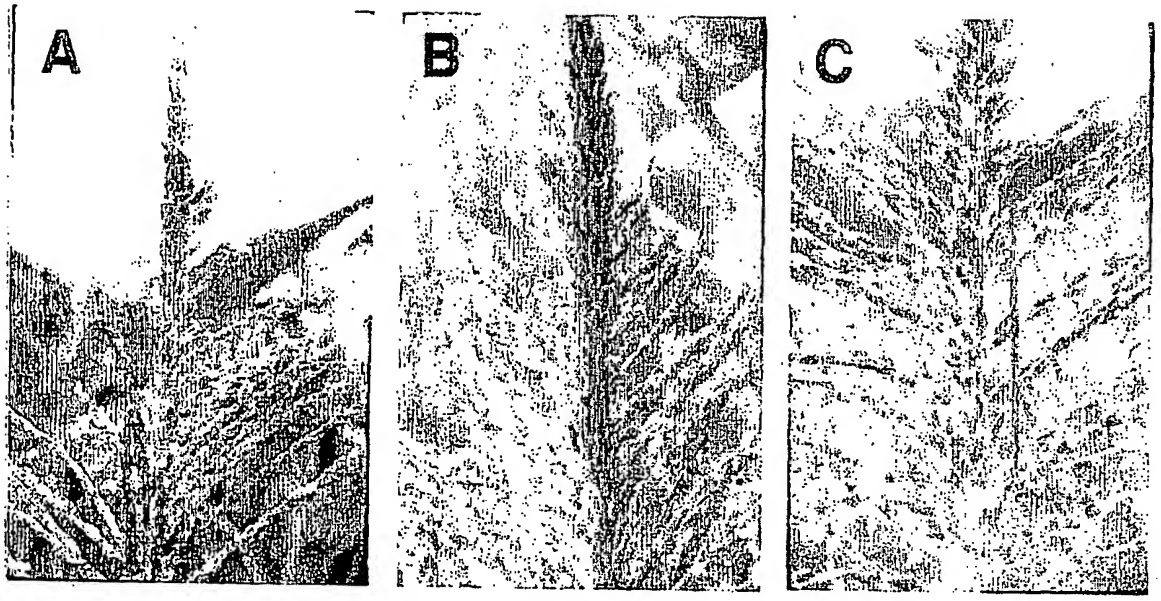

A
B
C

Fig. 5

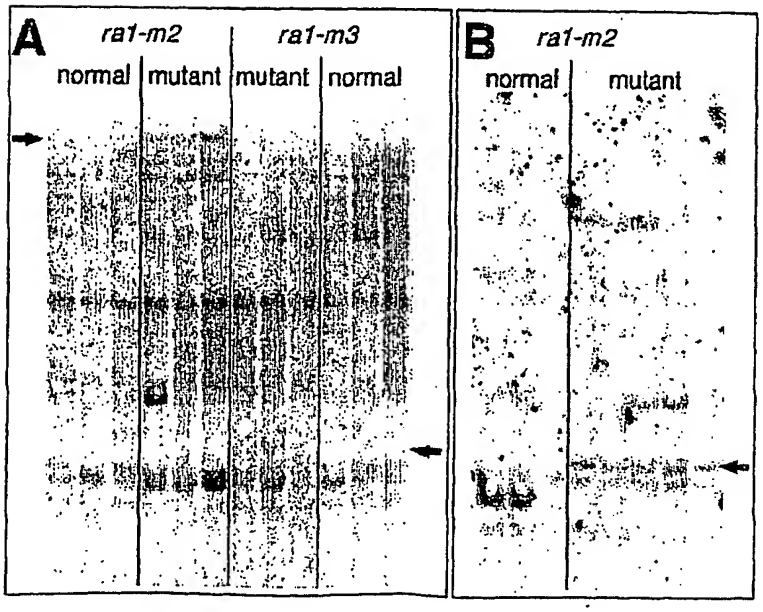

A
ra1-m2
ra1-m3
normal
mutant
mutant
normal
B
ra1-m2
normal
mutant

Fig. 6

6/21

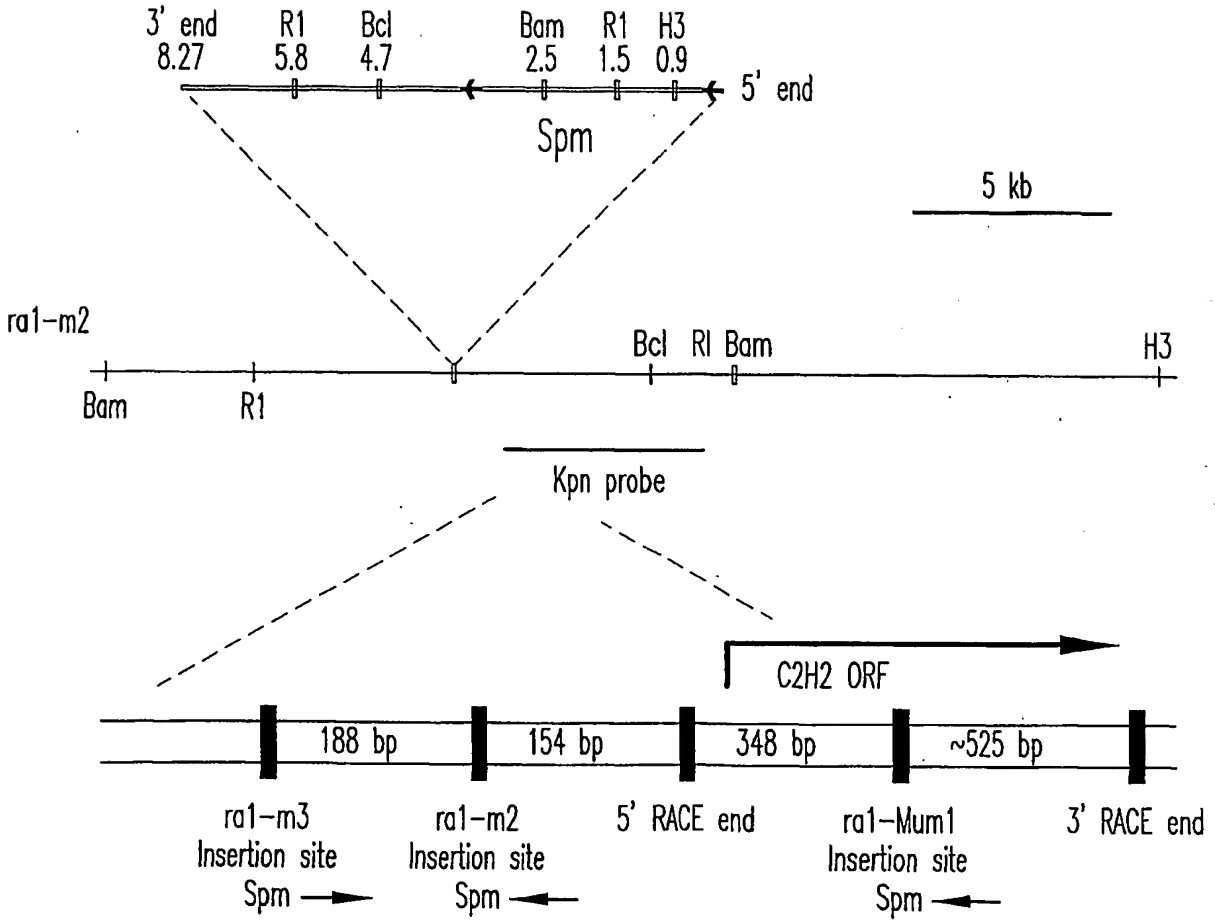

3' end
8.27
R1
5.8
Bcl
4.7
Bam
2.5
R1
1.5
H3
0.9
5' end
Spm
5 kb
ra1-m2
Bcl RI Bam
H3
Bam
R1
Kpn probe
C2H2 ORF
188 bp
154 bp
348 bp
~525 bp
ra1-m3
Insertion site
Spm
ra1-m2
Insertion site
Spm
5' RACE end
ra1-Mum1
Insertion site
Spm
3' RACE end

Fig. 7

```
                                                                  Score     E
Sequences producing significant alignments:                       (bits)  Value

pir||S60325  transcription factor SUPERMAN - Arabidopsis tha...     73  1e-12
gb|AAD23724.1|AC005956_13  (AC005956) putative SUPERMAN-like...    64  1e-09
pir||D71448  probable Zn finger protein - Arabidopsis thalia...    62  2e-09
pir||T02540  hypothetical protein F13M22.24 - Arabidopsis th...    61  9e-09
gb|AAF14043.1|AC011436_27  (AC011436) putative C2H2-type zin...    52  3e-06
pir||S55886  CCHH finger protein 6 - Arabidopsis thaliana >g...    52  3e-06
pir||S55882  CCHH finger protein 2 - Arabidopsis thaliana >g...    47  1e-04
gb|AAF26478.1|AC016447_1  (AC016447) putative zinc finger pr...    45  6e-04
pir||S55887  CCHH finger protein 7 - Arabidopsis thaliana >g...    44  8e-04
pir||S55881  CCHH finger protein 1 - Arabidopsis thaliana >g...    44  8e-04

 pir||S60325 transcription factor SUPERMAN - Arabidopsis thaliana
           >gi|1079669|gb|AAC49116.1| (U38946) SUPERMAN
           [Arabidopsis thaliana] >gi|1585427|prf||2124420A
           SUPERMAN gene [Arabidopsis thaliana]
           Length = 204

 Score = 73.4 bits (177), Expect = 1e-12
 Identities = 59/161 (36%), Positives = 79/161 (48%), Gaps = 34/161 (21%)

Query: 45  SYTCGYCKKEFRSAQGLGGHMNIHRLDRARLIHQQ--YTSHRIAAPHPNPNPSCTSVLD- 101
           SYTC +CK+EFRSAQ LGGHMN+HR DRARL  QQ   +S   + P+PNPN S +++ +
Sbjct: 46  SYTCSFCKREFRSAQALGGHMNVHRRDRARLRLQQSPSSSSTPSPPYPNPNYSYSTMANS 105

Query: 102 -------------------------LELSLSSLLAHGA--ASSDGGLSVPVAKLAGNRFS 134
                                   L  SLS    H    A       S+ V       RF+
Sbjct: 106 PPPHHSPLTLFPTLSPPSSPRYRAGLIRSLSPKSKHTPENACKTKKSSLLVEAGEATRFT 165

Query: 135 SASP-PTTKDVEGKNLELRIGACSHGDGAEERLDLQLRLGY 174
           S       ++ E  +LEL IG  +    +E+ LDL+LRLG+
Sbjct: 166 SKDACKILRNDEIISLELEIGLINE---SEQDLDLELRLGF 203
```

THE *RA1* CDNA SEQUENCE

```
                    20                  40                  60
CAAAGGTAGT TAGCTAGGTT AGGCACACGC GCGCCACTCG ACTAGCTAGC AGCTATGGAG
GTTTCCATCA ATCGATCCAA TCCGTGTGCG CGCGGTGAGC TGATCGATCG TCGATACCTC
                                                      M  E

                    80                 100                 120
GGAGAAGATG ACGGCGCCCA AATGAAACTG CAGCAACAAC AACAGTCGCC TTGCAGTGAC
CCTCTTCTAC TGCCGCGGGT TTACTTTGAC GTCGTTGTTG TTGTCAGCGG AACGTCACTG
 G  E  D   D  G  A  Q   M  K  L   Q  Q  Q   Q  Q  S  P   C  S  D

                   140                 160                 180
AACTTGAGCT TGTCCGCCGC CTCCTCATGG CTGCCGCCAC AGGTAAGGTC GTCGTCGTCG
TTGAACTCGA ACAGGCGGCG GAGGAGTACC GACGGCGGTG TCCATTCCAG CAGCAGCAGC
 N  L  S   L  S  A  A   S  S  W   L  P  P   Q  V  R  S   S  S  S

                   200                 220                 240
TCGTCGTCGT ACACCTGCGG GTATTGCAAG AAGGAGTTCA GATCAGCACA AGGGCTGGGA
AGCAGCAGCA TGTGGACGCC CATAACGTTC TTCCTCAAGT CTAGTCGTGT TCCCGACCCT
 S  S  S   Y  T  C  G   Y  C  K   K  E  F   R  S  A  Q   G  L  G

                   260                 280                 300
GGCCACATGA ACATCCACAG GCTGGACAGG GCCAGACTGA TCCACCAACA GTACACTTCA
CCGGTGTACT TGTAGGTGTC CGACCTGTCC CGGTCTGACT AGGTGGTTGT CATGTGAAGT
 G  H  M   N  I  H  R   L  D  R   A  R  L   I  H  Q  Q   Y  T  S

                   320                 340                 360
CACCGTATTG CTGCTCCCCA TCCAAACCCT AATCCTAGTT GCACATCAGT TCTTGACCTT
GTGGCATAAC GACGAGGGGT AGGTTTGGGA TTAGGATCAA CGTGTAGTCA AGAACTGGAA
 H  R  I   A  A  P  H   P  N  P   N  P  S   C  T  S  V   L  D  L

                   380                 400                 420
GAGCTCAGCT TGTCGTCGCT GCTAGCGCAT GGTGCTGCCA GCAGCGACGG AGGCTTGTCT
CTCGAGTCGA ACAGCAGCGA CGATCGCGTA CCACGACGGT CGTCGCTGCC TCCGAACAGA
 E  L  S   L  S  S  L   L  A  H   G  A  A   S  S  D  G   G  L  S

                   440                 460                 480
GTTCCAGTGG CAAAGCTGGC GGGCAACCGT TTCTCCTCCG CATCGCCCCC CACGACCAAG
CAAGGTCACC GTTTCGACCG CCCGTTGGCA AAGAGGAGGC GTAGCGGGGG GTGCTGGTTC
 V  P  V   A  K  L  A   G  N  R   F  S  S   A  S  P  P   T  T  K

                   500                 520                 540
GACGTCGAGG GGAAGAACTT AGAGTTGAGG ATAGGAGCGT GCAGTCATGG CGATGGCGCG
CTGCAGCTCC CCTTCTTGAA TCTCAACTCC TATCCTCGCA CGTCAGTACC GCTACCGCGC
 D  V  E   G  K  N  L   E  L  R   I  G  A   C  S  H  G   D  G  A

                   560                 580                 600
GAAGAGCGTC TGGATCTTCA GCTTAGACTG GGCTACTACT GAGCCAGACA GAGGAACGAA
CTTCTCGCAG ACCTAGAAGT CGAATCTGAC CCGATGATGA CTCGGTCTGT CTCCTTGCTT
 E  E  R.  L  D  L  Q   L  R  L   G  Y  Y   *

                   620                 640                 660
CTGCTACAAT GGGTACGTGC AGTGCATGAT GATGGAATGA CTGGCTTTGT ATAATAATAA
GACGATGTTA CCCATGCACG TCACGTACTA CTACCTTACT GACCGAAACA TATTATTATT

                   680                 700                 720
TGATGATCCG ATTATTGTTA TTTCTGTATG CTAAATATAT GTCTCTTATG TTAGATTTAA
ACTACTAGGC TAATAACAAT AAAGACATAC GATTTATATA CAGAGAATAC AATCTAAATT

TATAAAAAAA AAAA
ATATTTTTTT TTTT
```

Fig. 9

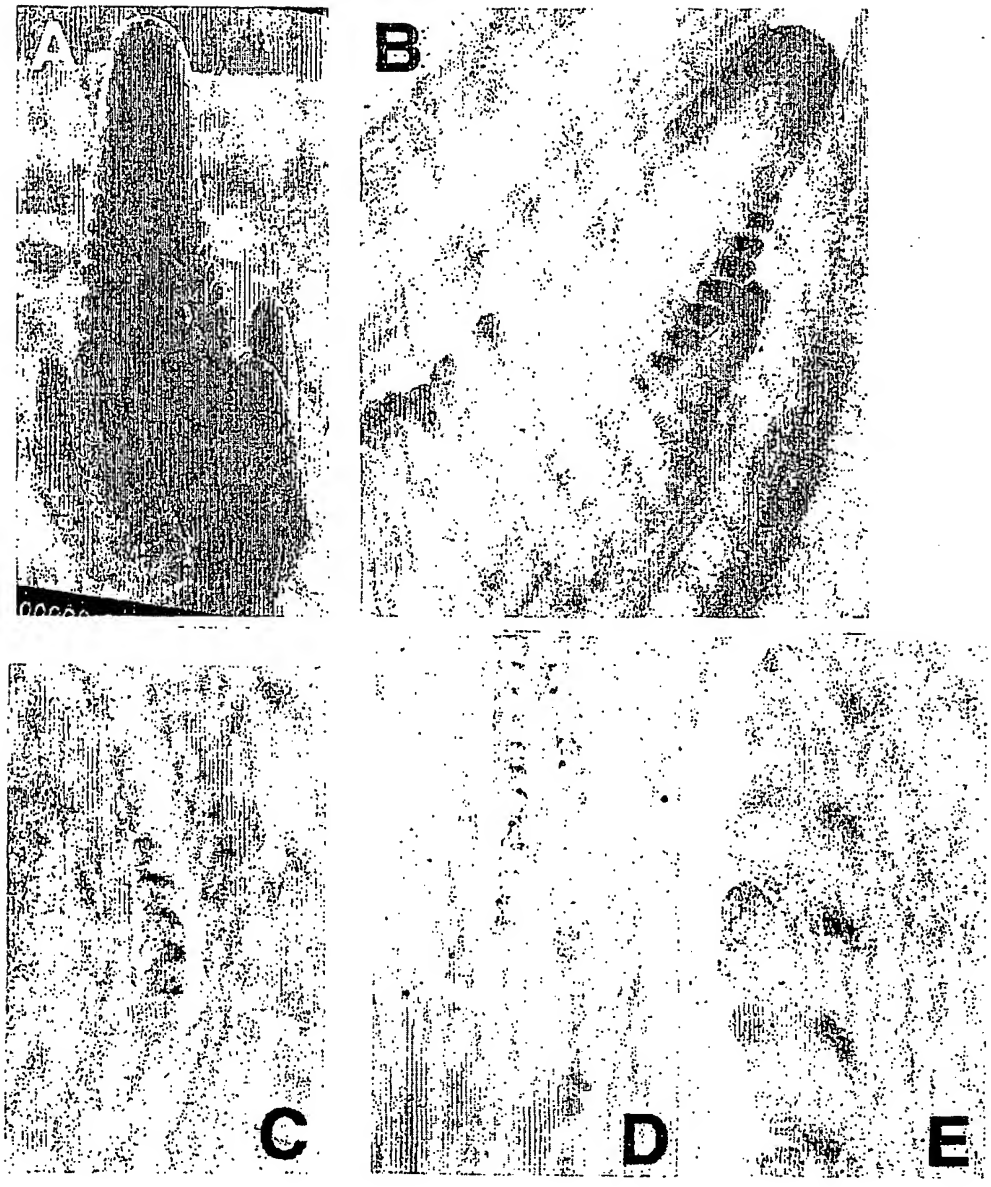


Fig. 10

```
                    20                  40                  60
AAGCTTACACAGTAACACGGGTGCCAATCTTTCCAAGTCTGCAACCACGGTCCGAGATAA
TTCGAATGTGTCATTGTGCCCACGGTTAGAAAGGTTCAGACGTTGGTGCCAGGCTCTATT

                    80                  100                 120
CTCTTTTGCACAAAGCTGATGGAAGAAATAGCTCAACTCTGAAAGCGCTAGCCAGACATG
GAGAAAACGTGTTTCGACTACCTTCTTTATCGAGTTGAGACTTTCGCGATCGGTCTGTAC

                   140                  160                 180
CTCAGGGACATAGCCTCGAACCATCGCAGGAAGAATCCGTTCAATCCATATGTGGAAGTC
GAGTCCCTGTATCGGAGCTTGGTAGCGTCCTTCTTAGGCAAGTTAGGTATACACCTTCAG

                   200                  220                 240
ATGACTCTTCATCCCTAAGACTCGCATAGTAGATAAGTTCACCCCCCTACTCAGGTTAGT
TACTGAGAAGTAGGGATTCTGAGCGTATCATCTATTCAAGTGGGGGGATGAGTCCAATCA

                   260                  280                 300
TGCATACCCATCAGGCAACATCAACATCTTGATCCACTGAAGTACTTCCTTCCTCTGGGC
ACGTATGGGTAGTCCGTTGTAGTTGTAGAACTAGGTGACTTCATGAAGGAAGGAGACCCG

                   320                  340                 360
CCTGCTTAGGACGAAATCGGCCTTAGGCCTTCTCCACGTCTTTCCGCCACTTGGCGGCTT
GGACGAATCCTGCTTTAGCCGGAATCCGGAAGAGGTGCAGAAAGGCGGTGAACCGCCGAA

                   380                  400                 420
CATCTCTAGTTTTGGTCTATCACATAACGCTGCCAGATCCACTCTTGACTTCACATTATC
GTAGAGATCAAAACCAGATAGTGTATTGCGACGGTCTAGGTGAGAACTGAAGTGTAATAG

                   440                  460                 480
CTTTGACTTATCAGGAATGTCCATAATTATTGCCCATAGTGCCTCGGCAACATTCTTTTC
GAAACTGAATAGTCCTTACAGGTATTAATAACGGGTATCACGGAGCCGTTGTAAGAAAAG
```

Fig. 11A

```
                    500                 520                 540
AGTGTGCATCACGTCAATGTTGTGTGGAAGGAGCAGGTCGTCATAATAGGGGAGCCGAGT
TCACACGTAGTGCAGTTACAACACACCTTCCTCGTCCAGCAGTATTATCCCCTCGGCTCA

                    560                 580                 600
CAAGCCAGACTTATGTGTCCACATATGCTGCTCACCATATCCCACAAAACCACCTTCTGT
GTTCGGTCTGAATACACAGGTGTATACGACGAGTGGTATAGGGTGTTTTGGTGGAAGACA

                    620                 640                 660
ATTGGCCACGAGCCCATCTATCTGTTGTCGAATTTCGGCACCAGTCATCGTTGCAGGTGG
TAACCGGTGCTCGGGTAGATAGACAACAGCTTAAAGCCGTGGTCAGTAGCAACGTCCACC

                    680                 700                 720
GCGGTCTGTCACTACGACACCTTTCGTAAAGTTCTTGATGTCTAGGCGGAATGGATGGTC
CGCCAGACAGTGATGCTGTGGAAAGCATTTCAAGAACTACAGATCCGCCTTACCTACCAG

                    740                 760                 780
AGCAGGAAGAAATTGTCGATGTTTATCGAAGGACGAATATTTGCCACCCTTTTTCAACCA
TCGTCCTTCTTTAACAGCTACAAATAGCTTCCTGCTTATAAACGGTGGGAAAAAGTTGGT

                    800                 820                 840
AATGAACCTGACAGCTTCCTTGCAAACTGGGCATGGGAACTTACCGTGAACACACCAGGC
TTACTTGGACTGTCGAAGGAACGTTTGACCCGTACCCTTGAATGGCACTTGTGTGGTCCG

                    860                 880                 900
ACAGAATAGCCCATGCGCCGATAAGTCATGCATGGAGTACTGGTACCAAACATGCATTCT
TGTCTTATCGGGTACGCGGCTATTCAGTACGTACCTCATGACCATGGTTTGTACGTAAGA

                    920                 940                 960
GAAGTTTGCCTTCGTAGCTCGGTCATACTTCCATCCCCCTTCCTCCCAGGCACGTACCAA
CTTCAAACGGAAGCATCGAGCCAGTATGAAGGTAGGGGGAAGGAGGGTCCGTGCATGGTT

                    980                1000                1020
TTCATCGATCAAAGGCTCCATATACACGCCTTTGCCGAGTGTAGGAAACATGTAGGTCTA
AAGTAGCTAGTTTCCGAGGTATATGTGCGGAAACGGCTCACATCCTTTGTACATCCAGAT

                   1040                1060                1080
CATAGTGTAGGAACATACCACAAAAAGTTTGGGAGACAAAATCAAAAAAATAAAAATATA
GTATCACATCCTTGTATGGTGTTTTTCAAACCCTCTGTTTTAGTTTTTTTATTTTTATAT

                   1100                1120                1140
CTTTGCCGAGTGTCTAGAGAAGACTCTAGTCCTTTGCCGAGTGCCCACTACTTGGCACTC
GAAACGGCTCACAGATCTCTTCTGAGATCAGGAAACGGCTCACGGGTGATGAACCGTGAG

                   1160                1180                1200
GGCAAAGAAGACTCTTTGCCGAGTGCCAACCCTCGGCTCTCGACAAAGACTGACGGCCGT
```

Fig. 11B

```
CCGTTTCTTCTGAGAAACGGCTCACGGTTGGGAGCCGAGAGCTGTTTCTGACTGCCGGCA

                  1220                  1240                  1260
CAGCTTTGGGACGGCCGCTGACGACCCTTTGCCGAGCGCCCCCTTTGCCGAGTGTTTGAC
GTCGAAACCCTGCCGGCGACTGCTGGGAAACGGCTCGCGGGGGAAACGGCTCACAAACTG

                  1280                  1300                  1320
ACTCGGCAAACATGTCTTTGCCGAGTGGGGTCCTGTGCCGAGTGTCCAGCACTCGGTAAA
TGAGCCGTTTGTACAGAAACGGCTCACCCCAGGACACGGCTCACAGGTCGTGAGCCATTT

                  1340                  1360                  1380
GAGGCTCGTTGCCGAGAGCCTAACTTTACCGAGTGCGGCTCTCGGCAAAGCCTTCTTTGC
CTCCGAGCAACGGCTCTCGGATTGAAATGGCTCACGCCGAGAGCCGTTTCGGAAGAAACG

                  1400                  1420                  1440
CGAGTGCCGGGGCACTCGGCAAAGAGGCCGACACTCGGCAAAGCCTCGGATTCTGGTAGT
GCTCACGGCCCCGTGAGCCGTTTCTCCGGCTGTGAGCCGTTTCGGAGCCTAAGACCATCA

                  1460                  1480                  1500
GAGAGCATGATTTAAGGGAAAAATTAAAACCAGTTTCATATGTCCGAGATAAAACATTTT
CTCTCGTACTAAATTCCCTTTTTAATTTTGGTCAAAGTATACAGGCTCTATTTTGTAAAA

                  1520                  1540                  1560
TTATCCCTTGCATCTGATCGGCCACATGTAGACTAAAATTGCCATCATCGAAATCGCTGA
AATAGGGAACGTAGACTAGCCGGTGTACATCTGATTTTAACGGTAGTAGCTTTAGCGACT

                  1580                  1600                  1620
TGTTGAAAAACAAATCCTAATTCCTATAAAACATCAGTTTCTCTAGTTTCCACAATCCTC
ACAACTTTTTGTTTAGGATTAAGGATATTTTGTAGTCAAAGAGATCAAAGGTGTTAGGAG

                  1640                  1660                  1680
CTTTTCTAAAAAAATTCATTATATTTCTTCTAATCTATCCCATCTTTATCTTTATAGTTT
GAAAAGATTTTTTTAAGTAATATAAAGAAGATTAGATAGGGTAGAAATAGAAATATCAAA

                  1700                  1720                  1740
CATTGTCGGGCGTCAAATCCGAAACAGGTAGATCTATTTTTTCATCACATTTGGCCTAAC
GTAACAGCCCGCAGTTTAGGCTTTGTCCATCTAGATAAAAAAGTAGTGTAAACCGGATTG

                  1760                  1780                  1800
AACCAGATACGCTTATCTCTTGTATGTGTAACCAAGCAACAACCACGAGCACTTCATCTC
TTGGTCTATGCGAATAGAGAACATACACATTGGTTCGTTGTTGGTGCTCGTGAAGTAGAG

                  1820                  1840                  1860
CAACTTTCCATCTATTTTCTTGTCTACGCCCTCTCTTACCCAGATTCTTCTACATCGCCA
GTTGAAAGGTAGATAAAAGAACAGATGCGGGAGAGAATGGGTCTAAGAAGATGTAGCGGT

                  1880                  1900                  1920
```

Fig. 11C

```
TAGAGGATGTAGAGGACCTCGAGCAGGCCATCTAGCTGGCCTCCGACTCGCCAGTCGTCG
ATCTCCTACATCTCCTGGAGCTCGTCCGGTAGATCGACCGGAGGCTGAGCGGTCAGCAGC

                  1940                1960                1980
CGGCGGCGGCGTAGATATATTTTTCTCCCCCACGTCTCCTCTAACAACCAGATAAGTCTT
GCCGCCGCCGCATCTATATAAAAAGAGGGGGTGCAGAGGAGATTGTTGGTCTATTCAGAA

                  2000                2020                2040
GTATATTGTATGTCGTGACTCTCCACCGCCATACAATACGTGGCTGAACAGCCAAGGGAG
CATATAACATACAGCACTGAGAGGTGGCGGTATGTTATGCACCGACTTGTCGGTTCCCTC

                  2060                2080                2100
AGAAAGAGGAGGCACCTATGACGTTCTCCTCCTTTATTTTGCGGTCCTTTATTAATCCCA
TCTTTCTCCTCCGTGGATACTGCAAGAGGAGGAAATAAAACGCCAGGAAATAATTAGGGT

                  2120                2140                2160
ACTTTTCTATTTCTTTTCTTGTTTTCCCCTTTCCCACCTAGATTCATCCTTGCAGTGTAG
TGAAAAGATAAAGAAAAGAACAAAAGGGGAAAGGGTGGATCTAAGTAGGAACGTCACATC

                  2180                2200                2220
ATCTATTTTTCTTCCACACGTCTAGCCTAACAACTAGATAAGATAAAACTTATATCTTAT
TAGATAAAAAGAAGGTGTGCAGATCGGATTGTTGATCTATTCTATTTTGAATATAGAATA

                  2240                2260                2280
ACTCTCTGTTCCAAATTAAAATTTGTTTTAGTGAATTATTGGATTCAAACAATTCTTGAT
TGAGAGACAAGGTTTAATTTTAAACAAAATCACTTAATAACCTAAGTTTGTTAAGAACTA

                  2300                2320                2340
ATTTTGTATATGTGTCTAGATTTATCATCATTTATTTGAATATATAGATAAAAAACAATA
TAAAACATATACACAGATCTAAATAGTAGTAAATAAACTTATATATCTATTTTTTGTTAT

                  2360                2380                2400
GTTAAAACGAATATTATTTTAAGACGGAGCGAGTATATCATCATACGATACGTGGCTGAT
CAATTTTGCTTATAATAAAATTCTGCCTCGCTCATATAGTAGTATGCTATGCACCGACTA

                  2420                2440                2460
CTCACAATCTCAACGTGGTCAAAGTTGTGTGTGCCGGGCCATCTGCGCGTCGTGTGACAC
GAGTGTTAGAGTTGCACCAGTTTCAACACACACGGCCCGGTAGACGCGCAGCACACTGTG

                  2480                2500                2520
CGGTGCATGCGCAGCCTTTTGTTTTGCCGCCCCGCCCGCTCCATGCATGGCATGGGTGCA
GCCACGTACGCGTCGGAAAACAAAACGGCGGGGCGGGCGAGGTACGTACCGTACCCACGT

                                    ral-m3::Spm
                  2540                2560                2580
GGTTCTGTAGCTATGCCCGGAAGCACCTAGCTAGCTCGCAGCCTACATCTGCAAACTCAC
CCAAGACATCGATACGGGCCTTCGTGGATCGATCGAGCGTCGGATGTAGACGTTTGAGTG
```

Fig. 11D

```
                2600                2620                2640
AAAGTTTGGGTATCGGAGGCATCAGCAGGTCGGGTTCAATGGAACGACGGATCACGTCTG
TTTCAAACCCATAGCCTCCGTAGTCGTCCAGCCCAAGTTACCTTGCTGCCTAGTGCAGAC

                2660                2680                2700
TGTGTCGCTTTCGCAGCAGCGGGGAGAGCGCGGGGCCCGGCCCAGGACGCATGGACCGAT
ACACAGCGAAAGCGTCGTCGCCCCTCTCGCGCCCCGGGCCGGGTCCTGCGTACCTGGCTA

                                                ra1-m2::Spm
                2720                2740                2760
GGACGCATGCAGACCATTTTTGTTTTTGTTTTTGTTTTTGTTTTTTTCCTGTCTAAAATG
CCTGCGTACGTCTGGTAAAAACAAAAACAAAAACAAAAACAAAAAAAGGACAGATTTTAC

                2780                2800                2820
TAGGTGTGCTCTATCTTGCCTCTTCATGCGATAATGTGTGTGTATATATATACATGCCCT
ATCCACACGAGATAGAACGGAGAAGTACGCTATTACACACACATATATATATGTACGGGA

                2840                2860                2880
TCACTCTTCTTATAGCTCGCTAGCCCAGCTTTAGTTTATAGCACTCTCTCACTCAGTAGT
AGTGAGAAGAATATCGAGCGATCGGGTCGAAATCAAATATCGTGAGAGAGTGAGTCATCA

                        ra1-m1::Spm
                2900        →       2920                2940
CAGCTCCCTCCATTTGTCCATTCTCCAAAGGTAGTTAGCTAGGTTAGGCACACGCGCGCC
GTCGAGGGAGGTAAACAGGTAAGAGGTTTCCATCAATCGATCCAATCCGTGTGCGCGCGG

                2960                2980                3000
ACTCGACTAGCTAGCAGCTATGGAGGGAGAAGATGACGGCGCCCAAATGAAACTGCAGCA
TGAGCTGATCGATCGTCGATACCTCCCTCTTCTACTGCCGCGGGTTTACTTTGACGTCGT

                3020                3040                3060
ACAACAACAGTCGCCTTGCAGTGACAACTTGAGCTTGTCCGCCGCCTCCTCATGGCTGCC
TGTTGTTGTCAGCGGAACGTCACTGTTGAACTCGAACAGGCGGCGGAGGAGTACCGACGG

                3080                3100                3120
GCCACAGGTAAGGTCGTCGTCGTCGTCGTCGTACACCTGCGGGTATTGCAAGAAGGA
CGGTGTCCATTCCAGCAGCAGCAGCAGCAGCAGCATGTGGACGCCCATAACGTTCTTCCT

                3140                3160                3180
GTTCAGATCAGCACAAGGGCTGGGAGGCCACATGAACATCCACAGGCTGGACAGGGCCAG
CAAGTCTAGTCGTGTTCCCGACCCTCCGGTGTACTTGTAGGTGTCCGACCTGTCCCGGTC

                3200                3220                3240
ACTGATCCACCAACAGTACACTTCACACCGTATTGCTGCTCCCCATCCAAACCCTAATCC
TGACTAGGTGGTTGTCATGTGAAGTGTGGCATAACGACGAGGGGTAGGTTTGGGATTAGG
```

Fig. 11E

ra1-m4(DI)::GT
ra1-m4::Spm *ra1-PI267184::GT* *ra1-LR::187bp*

```
                  3260                3280                3300
TAGTTGCACATCAGT^TCTTGACCTTGAGCTCAGCTTGT^CGTCGCTGCTAGC^GCATGGTGC
ATCAACGTGTAGTCA AGAACTGGAACTCGAGTCGAACA GCAGCGACGATCG CGTACCACG

                  3320                3340                3360
TGCCAGCAGCGACGGAGGCTTGTCTGTTCCAGTGGCAAAGCTGGCGGGCAACCGTTTCTC
ACGGTCGTCGCTGCCTCCGAACAGACAAGGTCACCGTTTCGACCGCCCGTTGGCAAAGAG

                  3380                3400                3420
CTCCGCATCGCTCCCCACGACCAAGGACGTCGAGGGGAAGAACTTAGAGTTGAGGATAGG
GAGGCGTAGCGAGGGGTGCTGGTTCCTGCAGCTCCCCTTCTTGAATCTCAACTCCTATCC

                  3440                3460                3480
AGCGTGCAGTCATGGCGATGGCGCGGAAGAGCGTCTGGATCTTCAGCTTAGACTGGGCTA
TCGCACGTCAGTACCGCTACCGCGCCTTCTCGCAGACCTAGAAGTCGAATCTGACCCGAT

ra1-63.3359::GTAC   ra1-PI267181/ra1-PI267495::[CACTA element]
                  3500                3520                3540
CTAC^TGAGCCAGACAGAGGAACGAACTGCTACAATGGGTACGTGCAGTGCATGATGATGG
GATG ACTCGGTCTGTCTCCTTGCTTGACGATGTTACCCATGCACGTCACGTACTACTACC

                                      *                     *
                  3560                3580                3600
AATGACTGGCTTTGTATAATAATAATGATGATCCGATTATTGTTATTTCTGTATGCTAAA
TTACTGACCGAAACATATTATTATTACTACTAGGCTAATAACAATAAAGACATACGATTT

                  3620                3640        *       3660
TATATGTCTCTTATGTTAGATTTAATATATATGACATATTTTATCTAACTAAATTAAATA
ATATACAGAGAATACAATCTAAATTATATATACTGTATAAAATAGATTGATTTAATTTAT

                  3680                3700                3720
AATTATATATAGGCGTCAACGTATTAAATACGTCTAGGGCATCGTAGTCTTTCCGAGGTG
TTAATATATATCCGCAGTTGCATAATTTATGCAGATCCCGTAGCATCAGAAAGGCTCCAC

                  3740                3760                3780
TCTTAACGTAGGAGGCTTTGGGGCCATCGGACCCTCCGGGCTCCGGAGCTTTCAACGCCT
AGAATTGCATCCTCCGAAACCCCGGTAGCCTGGGAGGCCCGAGGCCTCGAAAGTTGCGGA

                  3800                3820                3840
CAACGGCGTCGAGATCGACCCTCTCACTACTGAAGACGGCAGACAAAACAAAATATACTC
GTTGCCGCAGCTCTAGCTGGGAGAGTGATGACTTCTGCCGTCTGTTTTGTTTTATATGAG

                  3860                3880                3900
GGCAAAATATTTGTAGAGTGCCACACTGAGCAAGAGTACTCAATGAATCAAATGTCGGCA
CCGTTTTATAAACATCTCACGGTGTGACTCGTTCTCATGAGTTACTTAGTTTACAGCCGT
```

Fig. 11F

```
                3920                3940                3960
CATAGACAAGCTGACCGAAGTTAAATGACCGACAAAGATTTTTCGGTTTCTATTCTGTGA
GTATCTGTTCGACTGGCTTCAATTTACTGGCTGTTTCTAAAAAGCCAAAGATAAGACACT

                3980                4000                4020
AGTCCTAGGCTGGCATGTTGTCATATATTGAAGCTAGGAGCAAAGTTTAGCACTGCACGC
TCAGGATCCGACCGTACAACAGTATATAACTTCGATCCTCGTTTCAAATCGTGACGTGCG

                4040                4060                4080
AGACATGGTTGTCAATGTTGTGGGTGCACCTTGTAGTATTTAATACTCCCTCAAGGTGCA
TCTGTACCAACAGTTACAACACCCACGTGGAACATCATAAATTATGAGGGAGTTCCACGT

                4100                4120                4140
AATTATAAGTCGTTTAGGAAAACGACAGGTACTCCAAAATATAGCTTTGACCAATATTTT
TTAATATTCAGCAAATCCTTTTGCTGTCCATGAGGTTTTATATCGAAACTGGTTATAAAA

                4160                4180                4200
TTTTTAAATACAAATGAACTCTTAATACATTTATACTTTCGTAAAAGTACTTTTTAGGAT
AAAAATTTATGTTTACTTGAGAATTATGTAAATATGAAAGCATTTTCATGAAAAATCCTA

                4220                4240                4260
AAATCGACACATATGACTATTAGGTTTCAAAGCTAAATAACAAAACGATTATTTGTAGTC
TTTAGCTGTGTATACTGATAATCCAAAGTTTCGATTTATTGTTTTGCTAATAAACATCAG

                4280                4300                4320
AATATTTTACAAGTTTCATTTAATCCTTGTCCAGAACAACTTATAAGTTGGACAAGCTAG
TTATAAAATGTTCAAAGTAAATTAGGAACAGGTCTTGTTGAATATTCAACCTGTTCGATC

                4340                4360                4380
GCCTTGCAAATCCGATGGTTGTGGAAAGAAAAAATGATGTTATGAGGCCTTGGAAAGGAT
CGGAACGTTTAGGCTACCAACACCTTTCTTTTTTACTACAATACTCCGGAACCTTTCCTA

                4400                4420                4440
TGGAAATTCCTATCCTTCCCAACGCCCTTGCAATGGCAATAACACGTTGTTTTAGTCAGA
ACCTTTAAGGATAGGAAGGGTTGCGGGAACGTTACCGTTATTGTGCAACAAAATCAGTCT

                4460                4480                4500
TAAATGGATCAGAGGTTACAGAGTATCAGATTTAGCTCCTTCTCTTATGACTGTAGTTCT
ATTTACCTAGTCTCCAATGTCTCATAGTCTAAATCGAGGAAGAGAATACTGACATCAAGA

                4520                4540                4560
GAAAAAGATTAAGGGCTAGTTTTGGAACCCCATTTTCCCACGGGATTTTTATTTTTCCAA
CTTTTTCTAATTCCCGATCAAAACCTTGGGGTAAAAGGGTGCCCTAAAAATAAAAAGGTT

                4580                4600                4620
GGGAAATTAGTTTATTTTCCTTTGGGAAATATGAATCCCTTGTGAAAACGTAGTTCCTAA
```

Fig. 11C

```
CCCTTTAATCAAATAAAAGGAAACCCTTTATACTTAGGGAACACTTTTGCATCAAGGATT

                  4640                  4660                  4680
ACTAACCCTAAGAGTAAGAGATTGGTGATCGAGGCTCTAGAGGACAATCGATGGGTTAAT
TGATTGGGATTCTCATTCTCTAACCACTAGCTCCGAGATCTCCTGTTAGCTACCCAATTA

                  4700                  4720                  4740
GACATTAATGAAATTTATTCGTCGTCGAGCTTTATGAGTTTTTTCTTCTTTGGGATGTTG
CTGTAATTACTTTAAATAAGCAGCAGCTCGAAATACTCAAAAAAGAAGAAACCCTACAAC

                  4760                  4780                  4800
TCCAGGAAATCATTCTATTTGATCATGAGGACCAACATTTTTGGAAGCTCACAAGCTCAG
AGGTCCTTTAGTAAGATAAACTAGTACTCCTGGTTGTAAAAACCTTCGAGTGTTCGAGTC

                  4820                  4840                  4860
GAATCTACTCGCTTGATCAGCTTACTTGGCCTTTTTTTCAAAGGTCCCTAGCTTTTGAGC
CTTAGATGAGCGAACTAGTCGAATGAACCGGAAAAAAAGTTTCCAGGGATCGAAAACTCG

                  4880                  4900                  4920
ACGGGAAATGCATTTGGAAATCGTGTGCTCCCCCTAAATGCAAAATCTTCCTGAGCCTTG
TGCCCTTTACGTAAACCTTTAGCACACGAGGGGGATTTACGTTTTAGAAGGACTCGGAAC

TTGTTAGGAACAAATG
AACAATCCTTGTTTAC
```

ra1-LR has the following 187 base pair insertion:

```
                   20                    40                    60
AGAAGCGGGCCCAGACATTTGAGATTGGGTATTCAAAAATTTAAAAGATTAAAGAATTTA
TCTTCGCCCGGGTCTGTAAACTCTAACCCATAAGTTTTTAAATTTTCTAATTTCTTAAAT

                   80                   100                   120
GTGTTGTAACACTATTTTATGTAATACATTATTGACAAATTAATGTTCTAACACTATAGA
CACAACATTGTGATAAAATACATTATGTAATAACTGTTTAATTACAAGATTGTGATATCT

                  140                   160                   180
TTACCAAAAACATGGGTATTCAGTGAATACCCATGAAACCCCCCTGGGCCCGCCCATGGC
AATGGTTTTTGTACCCATAAGTCACTTATGGGTACTTTGGGGGGACCCGGGCGGGTACCG

TGCTAGC
ACGATCG
```

Fig. 11H

```
                    20                  40                  60
CAAAGGTAGTTAGCTAGGTTAGGCACACGCGCGCCACTCGACTAGCTAGCAGCTATGGAG
GTTTCCATCAATCGATCCAATCCGTGTGCGCGCGGTGAGCTGATCGATCGTCGATACCTC

                    80                  100                 120
GGAGAAGATGACGGCGCCCAAATGAAACTGCAGCAACAACAACAGTCGCCTTGCAGTGAC
CCTCTTCTACTGCCGCGGGTTTACTTTGACGTCGTTGTTGTTGTCAGCGGAACGTCACTG

                    140                 160                 180
AACTTGAGCTTGTCCGCCGCCTCCTCATGGCTGCCGCCACAGGTAAGGTCGTCGTCGTCG
TTGAACTCGAACAGGCGGCGGAGGAGTACCGACGGCGGTGTCCATTCCAGCAGCAGCAGC

                    200                 220                 240
TCGTCGTACACCTGCGGGTATTGCAAGAAGGAGTTCAGATCAGCACAAGGGCTGGAAGGC
AGCAGCATGTGGACGCCCATAACGTTCTTCCTCAAGTCTAGTCGTGTTCCCGACCTTCCG

                    260                 280                 300
AACATGAACATCCACAGGCTGGACAGGGCCAGACTGATCCACCAACAGTATACTTCACAC
TTGTACTTGTAGGTGTCCGACCTGTCCCGGTCTGACTAGGTGGTTGTCATATGAAGTGTG

                    320                 340                 360
CGTATTGCTACTCCCCATCCAAACCCTAATCCTAGTTGCACATCAGTTCTTGACCTTGAG
GCATAACGATGAGGGGTAGGTTTGGGATTAGGATCAACGTGTAGTCAAGAACTGGAACTC

                    380                 400                 420
CTCAGCTTGTCGTCGCTGCTAGCGCATGGTGCTGCCAGCAGCGACGGAGGCTTGTCTGTT
GAGTCGAACAGCAGCGACGATCGCGTACCACGACGGTCGTCGCTGCCTCCGAACAGACAA

                    440                 460                 480
CCAGTGGCAAAGCTGGCGGGCAACCGTTTCTCCTCCGCATCGCCCCACACGACCAAGGAC
GGTCACCGTTTCGACCGCCCGTTGGCAAAGAGGAGGCGTAGCGGGGTGTGCTGGTTCCTG

                    500                 520                 540
GTCGAGGGGAAGAACTTAGAGTTGAGGATAGGAGCGTGCAGTCATGGCGATGGCGCGGAA
CAGCTCCCCTTCTTGAATCTCAACTCCTATCCTCGCACGTCAGTACCGCTACCGCGCCTT

                    560                 580                 600
GAGCGTCTGGATCTTCAGCTTAGACTGGGCTACTACTGAGCCAGACAGAGGAACGAACTG
CTCGCAGACCTAGAAGTCGAATCTGACCCGATGATGACTCGGTCTGTCTCCTTGCTTGAC

                    620                 640                 660
CTACAATGGGTACGTGCAGTGCATGATGATGGAATGACTGGCTTTGTATAATAATAATGA
GATGTTACCCATGCACGTCACGTACTACTACCTTACTGACCGAAACATATTATTATTACT

                    680                 700
TGATCCGATTATTGTTATTTCTGTATGCTAAAAAAAAAAAAAAAAA
ACTAGGCTAATAACAATAAAGACATACGATTTTTTTTTTTTTTTT
```

Fig. 12

```
                    20                  40                  60
AGAAGCGGGCCCAGACATTTGAGATTGGGTATTCAAAAATTTAAAAGATTAAAGAATTTA
TCTTCGCCCGGGTCTGTAAACTCTAACCCATAAGTTTTTAAATTTTCTAATTTCTTAAAT

                    80                  100                 120
GTGTTGTAACACTATTTTATGTAATACATTATTGACAAATTAATGTTCTAACACTATAGA
CACAACATTGTGATAAAATACATTATGTAATAACTGTTTAATTACAAGATTGTGATATCT

                   140                  160                 180
TTACCAAAAACATGGGTATTCAGTGAATACCCATGAAACCCCCCTGGGCCCGCCCATGGC
AATGGTTTTTGTACCCATAAGTCACTTATGGGTACTTTGGGGGGACCCGGGCGGGTACCG

TGCTAGC
ACGATCG
```

Fig. 13

```
                    20                  40                  60
TCACAAGTTTGGGTATCGGAGGCATCAGCAGGTCGGGTTCAATGGAACGACGGATCACGT
AGTGTTCAAACCCATAGCCTCCGTAGTCGTCCAGCCCAAGTTACCTTGCTGCCTAGTGCA

                    80                  100                 120
CTGTGTGTCGCTTTCGCAGCAGCGGGGAGAGCGCGGGGCCCGGCCCAGGACGCATGGACC
GACACACAGCGAAAGCGTCGTCGCCCCTCTCGCGCCCCGGGCCGGGTCCTGCGTACCTGG

                   140                  160                 180
GATGGACGCATGCAGACCATTTTTGTTTTTGTTTTTGTTTTTTTCCTGTCTAAAATGTAG
CTACCTGCGTACGTCTGGTAAAAACAAAAACAAAAACAAAAAAAGGACAGATTTTACATC

                   200                  220                 240
AAACTGTGCATGTGTGCAATGTGTGCTCTATCTTGCCTCTTCATGCGGATGATGTGTGTA
TTTGACACGTACACACGTTACACACGAGATAGAACGGAGAAGTACGCCTACTACACACAT

                   260                  280                 300
TATATATACATGCCCTTCACTCTTCTTAGCTCGCTAGCCCAGCTTTAGTTTATAGCACTC
ATATATATGTACGGGAAGTGAGAAGAATCGAGCGATCGGGTCGAAATCAAATATCGTGAG

                   320                  340                 360
TCTCACTCAGTAGTCAGCTCCCTCCATTTATCCATTCTCCAAAGGTAGTTAGCTAGGTTA
AGAGTGAGTCATCAGTCGAGGGAGGTAAATAGGTAAGAGGTTTCCATCAATCGATCCAAT

                   380                  400                 420
GGCACACGCGCGCCACTCGACTAGCTAGCAGCTATGGAGGGAGAAGATGACGGCGCCCAA
CCGTGTGCGCGCGGTGAGCTGATCGATCGTCGATACCTCCCTCTTCTACTGCCGCGGGTT

                   440                  460                 480
ATGAAACTGCAGCAACAACAACAGTCGCCTTGCAGTGACAACTTGAGCTTGTCCGCCGCC
TACTTTGACGTCGTTGTTGTTGTCAGCGGAACGTCACTGTTGAACTCGAACAGGCGGCGG

                   500                  520                 540
TCCTCATGGCTGCCGCCACAGGTAAGGTCGTCGTCGTCGTCGTACACCTGCGGGTATTGC
AGGAGTACCGACGGCGGTGTCCATTCCAGCAGCAGCAGCAGCATGTGGACGCCCATAACG

                   560                  580                 600
AAGAAGGAGTTCAGATCAGCACAAGGGCTGGGAGGCCACATGAACATCCACAGGCTGGAC
TTCTTCCTCAAGTCTAGTCGTGTTCCCGACCCTCCGGTGTACTTGTAGGTGTCCGACCTG

                   620                  640                 660
AGGGCCAGACTGATCCACCAACAGTACACTTCACACCGTATTGCTGCTCCCCATCCAAAC
TCCCGGTCTGACTAGGTGGTTGTCATGTGAAGTGTGGCATAACGACGAGGGGTAGGTTTG

                   680                  700                 720
```

Fig. 14A

```
CCTAATCCTAGTTGCACATCAGTTCTTGACCTTGAGCTCAGCTTGTCGTCGCTGCTAGCG
GGATTAGGATCAACGTGTAGTCAAGAACTGGAACTCGAGTCGAACAGCAGCGACGATCGC

                  740                 760                 780
CACGGTGCTGCCAGCAGCGACGGAGGCTTGTCTGTTCCAGTGGCAAAGCTGGCGGGCAAC
GTGCCACGACGGTCGTCGCTGCCTCCGAACAGACAAGGTCACCGTTTCGACCGCCCGTTG

                  800                 820                 840
CGTTTCTCCTCCGCATCGCCCCCCACGACCAAGGACATCGAGGGGAAGAACTTAGAGTTG
GCAAAGAGGAGGCGTAGCGGGGGGTGCTGGTTCCTGTAGCTCCCCTTCTTGAATCTCAAC

                  860                 880                 900
AGGATAGGAGCGTGCAGTCATGGCGATGGCGCGGAAGAGCGTCTGGATCTTCAGCTTAGA
TCCTATCCTCGCACGTCAGTACCGCTACCGCGCCTTCTCGCAGACCTAGAAGTCGAATCT

                  920                 940                 960
CTGGGCTACTACTGAGCCAGACAGAGGAACGAACTGCTTCAATGGGTACGTGCAGTGCAT
GACCCGATGATGACTCGGTCTGTCTCCTTGCTTGACGAAGTTACCCATGCACGTCACGTA

                  980                1000                1020
GATGATGGAATGACTGGCTTTGTATAATAATAATGATGATCCGAATATTGTTATTTCTGT
CTACTACCTTACTGACCGAAACATATTATTATTACTACTAGGCTTATAACAATAAAGACA

                 1040                1060                1080
ATGCTAAATATATGTCTCTTATGTTAGATTTAATATATATGACTTATATTTTATCTAACT
TACGATTTATATACAGAGAATACAATCTAAATTATATATACTGAATATAAAATAGATTGA

                 1100                1120                1140
AAATTAAATAAATTATATATAGGCGTCAACGTATTAAATACGTCTAGGGCATCGTAGTCT
TTTAATTTATTTAATATATATCCGCAGTTGCATAATTTATGCAGATCCCGTAGCATCAGA

                 1160                1180                1200
TTCCGAGGTGTCTTAACGTAGGAGGCTTTGGGGCCATCGGACCCTCCGGGCTCCGGAGCT
AAGGCTCCACAGAATTGCATCCTCCGAAACCCCGGTAGCCTGGGAGGCCCGAGGCCTCGA

                 1220                1240                1260
TTCAACGCCTCAACGGCGTCGAGACCCTCTCACTACTGAAGACGGCAGACAAAACAAAAT
AAGTTGCGGAGTTGCCGCAGCTCTGGGAGAGTGATGACTTCTGCCGTCTGTTTTGTTTTA

                 1280                1300                1320
ATACTCGGCAAAATATTTGTAGAGTGCCACACTGAGCAAGAGTACTCAATGAATCAAATG
TATGAGCCGTTTTATAAACATCTCACGGTGTGACTCGTTCTCATGAGTTACTTAGTTTAC

                 1340                1360
TCGGCACATAGACAAGCTGACCGAAGTTAAATGACCGACAAAGATTTTTCGGTT
AGCCGTGTATCTGTTCGACTGGCTTCAATTTACTGGCTGTTTCTAAAAAGCCAA
```

Fig. 14B